DOCUMENT RESUME

ED 261 533 FL 015 121

AUTHOR Partee, Barbara H., Ed.; And Others
TITLE Workshop on Information and Representation Report.
PUB DATE Apr 85
NOTE 151p.; Workshop on Information and Representation (Washington, DC, March 30-April 1, 1985).
PUB TYPE Collected Works - Conference Proceedings (021) -- Reports - Descriptive (141)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS College Faculty; Communications; *Computational Linguistics; Federal Aid; *Financial Support; *Fund Raising; Graduate Study; *Information Processing; *Interdisciplinary Approach; Interprofessional Relationship; Needs Assessment; Networks; Research and Development Centers; Researchers; *Research Projects; Teacher Exchange Programs

ABSTRACT

The proceedings of a workshop on the study of information, computation, and cognition, a field of interdisciplinary research that includes communication research in artificial intelligence, computer science, linguistics, logic, philosophy, and psychology, gives an overview of the status of funding support for the field and the concerns of researchers for the continuation of research opportunities. Recommendations concerning fund raising that emerged from the workshop are presented, and papers providing background information are included. These papers treat the following subjects: interdisciplinary projects and topics; needs assessment; a description of the research community; an estimate of currently available resources; the history of the Alfred P. Sloan Foundation and System Development Foundation support; and the role of industrial research centers. Four expository papers are also included: "Language, Mind and Information" (John Perry); "A Presentation to the NSF Workshop on Information and Representation" (Jerry Fodor); "Research Goals in Information, Computation and Cognition" (Barbara Partee); and "A Presentation at NSF Workshop on Information and Representation" (Mitch Marcus). (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED261533

BEST COPY AVAILABLE

Report of Workshop on Information and Representation

Barbara H. Partee, Chair

Washington, D.C., March 30 through April 1, 1985

Editors

Barbara H. Partee, Stanley Peters, Richmond Thomason

Panelists

Stephen Anderson
Jaime Carbonell
Herbert Clark
David Evans
Jerry Fodor
Donald Foss
Lila Gleitman
Clark Glymour
Barbara Grosz
Aravind Joshi
Hans Kamp
Samuel Jay Keyser
Mitch Marcus
David MacQueen
Gary Olson
Barbara Partee
John Perry
Stanley Peters
Stanley Petrick
Robert Ritchie
Stanley Rosenschein
Dana Scott
Richmond Thomason
Thomas Wasow

Under the sponsorship of

National Science Foundation
System Development Foundation

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Foundations.

A copy of this report may be obtained by writing to Professor Stanley Peters, Department of Linguistics, Stanford University, Stanford, CA 94305. Permission is hereby granted to reproduce this report.

April 1985

FL015121

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Richmond H. Thomason

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Executive Summary

The study of Information, Computation and Cognition is a rapidly advancing area of science, in which progress has accelerated since the late 1970s. Led by basic research in The United States, this field is not only transforming concepts of computation but serves also as a vital basis for the development of intelligent machines and communications systems.

Seeking solutions to problems central to their own disciplines, computer scientists, linguists, logicians, philosophers, psychologists and workers in artificial intelligence have turned to one another; the results of these interactions will prove fundamental, we believe, to the development of the information processing systems of the future. With the support of large-scale seed money from a few private foundations, interdisciplinary theories are being developed that are scientifically exciting as well as technologically useful. Without timely public support, however, this scientific work cannot continue at a level that meets the urgent need.

A crisis is developing in funding for the necessary basic research. This crisis has several dimensions.

- Several of the disciplines involved are supported at very low levels by federal agencies.
- Private foundations that have seeded research in the field, with very striking results, are completing their grant programs.
- The corporations that wish to apply the results of this work cannot commercially justify large-scale support for high-risk basic research.

The level of long-term funding to Information, Computation and Cognition is insufficient to provide the stable environment required for basic research. Presently available federal funds are inadequate to provide researchers the following essential support:

- the computers and networking needed on a national basis,
- adequate support for research centers and project groups, which are vital supplements to individual research grants in this highly interdisciplinary field,
- training of the next generation of researchers in the field: graduate students, postdoctoral scholars, and senior research scientists.

Unless these financial problems are solved, the U.S. runs a serious risk of losing the momentum which has produced its lead in this important area of basic science and of failing to build a vital theoretical base for future technological development. The money needed is modest in comparison

with what is now committed to technology. The need to preserve the momentum of relevant scientific research is urgent. We call on the National Science Foundation to consider an initiative that will provide appropriate support for interdisciplinary basic research in Information, Computation and Cognition.

Recommendations

The panel recommends a national initiative to provide directed federal funding for basic research in Information, Computation and Cognition. Beginning in Fiscal Year 1987, a productive effort could be supported with $20 million annually in new funds, expanding each year over a five year period to twice that amount. Peer review of proposals for grants of these funds needs to be overseen by an appropriate interdisciplinary body.

The added funds should support

- acquisition and operation of computing equipment for basic research in this area,
- access to computer networks for researchers in the field around the nation,
- a spectrum of research activities, including reasonable funding for at least two major research centers and grants to collaborative group projects and to individual investigators,
- graduate and postdoctoral programs, the latter perhaps in part nationally administered,
- provision for greater exchange not only of faculty across disciplines and institutions but also of researchers from nonuniversity settings,
- increased funding for the National Science Foundation's programs in the contributing disciplines.

Table of Contents

Executive Summary / i
*Richmond Thomason**

Overview and Recommendations / 1
*Barbara Partee, Stanley Peters, Richmond Thomason**

History of the Report / 14
*Barbara Partee, Stanley Peters, Richmond Thomason**

Examples of Interdisciplinary Research Projects and Topics / 19
*Elizabeth Macken (Editor)***

Estimate of Needs / 63
*Jerry Fodor, Lila Gleitman, Jay Keyser, Stanley Peters, Robert Ritchie**

Description of Research Community / 72
*Thomas Wasow**

Estimate of Currently Available Resources / 74
*Richmond Thomason, Thomas Wasow**

History of Alfred P. Sloan Foundation and System Development Foundation Support / 78
Eric Wanner, Carl York

Role of Industrial Research Centers / 83
*David MacQueen, Mitch Marcus, Stanley Petrick, Robert Ritchie**

Expository Papers / 85

I. Language, Mind and Information / 87
John Perry

II. A Presentation to the NSF Workshop on Information and Representation / 106
Jerry Fodor

III. Research Goals in Information, Computation and Cognition / 118
Barbara Partee

IV. A Presentation at NSF Workshop on Information and Representation / 125
Mitch Marcus

---

* This section has been circulated to all workshop participants for comments. The authors named drafted the initial version and, except for Lila Gleitman who was out of the country, revised it in light of comments from other workshop participants.

** See preface of this section for further information about its preparation.

Overview and Recommendations

Introduction

With the advent of writing and calculation, human culture was utterly transformed. In our time, the computer has not only introduced a far more powerful means of preserving and processing information, but has suggested to cognitive scientists new paradigms for understanding thought and language.

The ensuing explosive growth of research on the acquisition, representation, and communication of information, by both humans and machines, comes as no surprise. More striking is its simultaneous, independent development within the boundaries of many separate disciplines—including artificial intelligence, computer science, linguistics, logic, philosophy, and psychology—followed by a mutual recognition of the need to transcend these disciplinary boundaries.

We begin our overview of the field of interdisciplinary research that is now emerging by briefly describing its focus and explaining its origins. We then lead up to our recommendations concerning its future support by reviewing the current state of support, sketching some mechanisms for basic research, estimating the nature and amount of support needs, comparing these with current trends, and discussing technological opportunities this field presents. We conclude with our recommendations for the support of basic research.

The New Field

Increasing recognition of shared problems has brought about a fruitful collaboration on many fronts, and led to an interdisciplinary pattern of research that is emerging in many universities and research centers across the country. Essential funding for this interdisciplinary research area was provided by the Sloan Foundation and the System Development Foundation, which together invested $46 million in nurturing the emerging field.

Because of the many contributing disciplines and the breadth of the subject matter, this research is expressing itself in a variety of different ways. The field is therefore best defined by the research problems that created and continue to animate it. In the section titled "Examples of Interdisciplinary Research Projects and Topics" (p. 19), we describe two dozen examples of such problems on topics such as The Structure of Human and Machine Representational Systems, Discourse and Problems of Context, Learning and Reasoning in Humans and Machines, and The Nature of Informational

Content. These examples range from completed research, which in some cases already has inspired applications in information technology, to entirely new projects. The "Examples" section, we hope, helps to communicate the diversity and excitement of this field, as well as its past accomplishments, its present directions, and its future promise.

In this document, we have decided to refer to this research area as *Information, Computation and Cognition*. Conscious of the potential importance of this report for the future, we have given careful thought to the question of nomenclature, despite the limitations of time under which we have labored. Our choice is a compromise intended to satisfy a number of different constraints, some more technical, some less so. It should be taken only as a provisional name; a shorter or more appropriate choice may well emerge as the area coalesces and matures.

At present, research in this field is carried out under many banners (for example, "Cognitive Science," "Language and Information," and "Applied Logic"). Often it is carried out anonymously, within a particular established discipline or a combined discipline, such as Computational Linguistics.

The lack of terminological uniformity that prevails is not a reflection of an intellectual confusion; such situations are typical when major realignments of disciplines are taking place. The fact that different groups have chosen (at least, as things stand now) to describe their efforts by somewhat different names reflects local color and different emphases, due to reasons that are largely historical.

This situation is very similar to the early days of Computer Science, when a multiplicity of names were coined to describe apparently the same activity. Even now, 20 years after the time that formal academic programs were set up, the name "Computer Science" is not uniformly accepted, although the *curriculum* of the subject has become quite stable and uniform, at the introductory and intermediate levels.

Research Background

Questions central to all of the core disciplines of Information, Computation and Cognition concern *information*, *representation*, *semantic content*, *mental states*, and *reasoning*. Linguistics and psychology are concerned with these issues as they arise in humans; computer science as they arise in machines; and philosophy, logic and artificial intelligence as they arise in humans and machines, and in general. Such studies, already interdisciplinary, are coming to share an emphasis on understanding the architecture and dynamics of finite, resource-limited devices which operate with partial and fallible information in changing contexts, and whose information processing is thoroughly connected to planning and action.

The following elements illustrate the rapid development of this research.

- The Chomskyan revolution in linguistics, launched in 1957, with its emphasis on the study of the human language faculty as a source of insight into human cognitive capacities.
- The related shift in psychology from behaviorism to a cognitive orientation.
- The birth and rapid growth of the field of artificial intelligence, starting from work in the late 1950s, such as Newell and Simon's General Problem Solver
- The integration of the tools and techniques of formal logic into the study of the semantics of both natural languages and computer languages. These developments originate with Montague's work on English as a formal language and Scott and Strachey's joint research on the semantics of programming languages, and their effects on the scientific community date from the early 1970s.
- The development during the 1970s, primarily from origins in ordinary language philosophy, of a systematic approach to pragmatics, or the study of language use.

The intrinsically interdisciplinary field of *cognitive science* was one major outgrowth of these developments; it came into focus partly through the intensive efforts, starting in 1977, of the Sloan Foundation to foster increased communication and collaboration across the boundaries of the component cognitive sciences, which had remained quite separate and sometimes hostile in the mid-seventies, and to help a single field of cognitive science coalesce conceptually and begin to make major progress on the very difficult problems posed by the analysis of intelligence. (See the statement in this report by Eric Wanner on the Sloan Program in Cognitive Science, p. 78.)

The more recent infusion of support from the System Development Foundation for basic research in related areas was targeted at similar research problems, but with an added emphasis on integrating foundational research on language, information, and computation from both cognitive and machine-oriented perspectives, and on seeking mathematically formalized theories that might elucidate potential commonalities between humans and computers in the intelligent acquisition and use of information through language and other means. (See the statement by Carl York on the System Development Foundation's Program in Computational Linguistics and Speech, p. 79.)

To many researchers in these areas, the differences between humans and computers are seen as being as important as the potential commonalities. The relation between minds and machines is still a highly controversial

issue (see, for instance, the expository paper by Fodor in this report), and of course not everyone interested in one topic is directly interested in the other. However, for some researchers the study of minds and machines is a single topic. And researchers from different disciplinary backgrounds and different schools of thought are increasingly recognizing that the need for interdisciplinary collaboration on these issues, and the benefits of such collaboration, transcend the current differences.

In order to make progress on the questions most central to their respective concerns, the core disciplines have traditionally idealized along certain dimensions. These idealizations are in many ways complementary. Pure logic, for example, has traditionally ignored resource limitations, while computer science has traditionally restricted its attention to databases describable within simple fragments of first-order extensional logic. Formal semanticists have developed powerful tools for representing meanings that unfortunately appear to be computationally intractable, while information processing models have typically posited symbolic representations and formal symbolic manipulations without providing a semantic account that connects these representations with their intended content.

Progress within the constraints of these idealizations has arrived a critical point. It is now (a) intellectually *possible*, in view of the contributions of neighboring disciplines, to replace current idealizations with more realistic assumptions, and (b) theoretically *necessary* to do so, not only in interdisciplinary research, but in order to overcome fundamental problems within the core disciplines themselves. The obstacles to seriously interdisciplinary collaboration on deep central issues are formidable, but we are convinced that such collaboration is the only way to forestall costly cycles of misunderstandings and incompatible idealizations which must be undone in later attempts to bridge the gaps between separate disciplinary approaches to what we can clearly identify as common problems.

At the present moment, we are only beginning to see the impact that these changes are effecting on their parent disciplines. Many of the researchers now pursuing *interdisciplinary* topics are leaders in their original disciplines; this enables the innovations they propose to attract considerable attention. As the parent disciplines begin to change in response to these pressures, new theories are created which are more applicable in related disciplines. This encourages further interactions, and the potential for theoretical growth is rapidly accelerated.

The progress we now see will contribute to theories in many scientific areas. Moreover, the benefits transcend the needs of pure science. Our present information technology rests on a number of stock ideas from the core disciplines, largely dating from many years ago. There is a widespread

feeling that the usefulness of these ideas is approaching its limit, as well as a growing sense of frustration at the amount of product development time that goes into rewriting old programs for slightly new purposes, and at the lack of criteria that can serve to compare and evaluate these programs.

These needs can only be filled by the generality that comes from relevant theory, and it is fortunate that such theories are developing now in a common effort to solve theoretical problems. We have already begun to see the usefulness of such scientific work; but to create an environment in which these developments can continue and be made available to future technology, there is an urgent need for new forms of support.

Present State of Support

The present, rapidly growing level of basic research activity would not have been possible without $46 million of support that the Sloan and System Development Foundations provided over a ten-year period. (The foundations' programs are described in a later section of this report.) Currently they give more than $10 million per year for interdisciplinary basic research in Information, Computation and Cognition, an increase from approximately $1 million in 1977, the first year of the Sloan Foundation's cognitive science program.

Federal funding of basic research in this area is estimated at approximately $8 million per year, largely from the National Science Foundation; other federal agencies support very little basic research in this field. (See the Estimate of Currently Available Resources in this report.) The private foundations have supported a variety of research groups and centers. Federal funding for basic research in the field goes entirely to individual researchers or very small combinations. It is mainly awarded through disciplinary programs; NSF's interdisciplinary program in Information Science is a notable exception. Adequate peer review of interdisciplinary proposals is a perennial problem.

Private industry gives some support to the work. This takes two forms: research in corporate research laboratories, and provision of equipment to academic researchers. There is one case of a corporation participating directly in a university research center, by assigning its own employees to conduct research there. All forms of support from industry for basic research in the field account for only a small part of the overall level, however. As the time from basic discovery to technological application may average ten years, it is not surprising that industry considers basic research too high in risk to justify a large investment of resources.

The critical importance private foundation funding has had for basic research in Information, Computation and Cognition is evident both from

the significant fraction of total support which it constitutes and also from the pervasive role it has played in supporting the work described in the "Examples" section of this report. The level of private and federal support has not been sufficient to meet the needs of basic research as the field has grown. However, a far more serious problem will present itself in the near future, when both the Sloan and the System Development Foundation programs for this field come to an end in approximately two years. At that point, enormous momentum for research progress will be lost unless federal support for basic interdisciplinary research in this area takes a quantum leap in magnitude.

Importance of Strong Core Disciplines

We shall recommend a major federal initiative for the funding of basic research in Information, Computation and Cognition, importantly including funding for research centers and project groups. We wish to stress that our recommendation is for new funds to be allocated in support of this interdisciplinary area. We do not recommend a redirection of funds from current federal programs supporting the core disciplines of this field. Several of these disciplinary programs are presently underfunded, and research in Information, Computation and Cognition would not benefit from their suffering. In fact, the cooperative research we discuss in this report would have been impossible without flourishing core disciplines, and would soon wither without them. Progress in interdisciplinary research remains crucially dependent on continued research support for these disciplines.

Necessary Research Environment

The inherently interdisciplinary nature of basic research in Information, Computation and Cognition contributes to certain of the field's significant characteristics. A thoroughly stable research environment is needed in order for it to flourish, because the intrinsic difficulty of bringing together theories that in many cases have developed with the aid of incompatible idealizations is increased by the problems of communicating across disciplinary lines. Long periods of concentrated research are necessary to overcome these obstacles if satisfying results such as the "Examples" section of this report describes for some problems are to be achieved. Uncertainties about support and the frequent necessity to seek new funding seriously degrade the research environment. One of the crucial needs of basic research in the field is for long-term funding, on the order of three- to five-years per grant.

A further requirement is the opportunity for researchers to interact easily with scientists from other disciplines who share an interest in fun-

damental problems of Information, Computation and Cognition. Several mechanisms are necessary to adequately meet this need. For one, nationwide networking of researchers in the field, regardless of their institution or home discipline, is an important priority. The need for computer networking is possibly most crucial for individuals who have few colleagues in the interdisciplinary endeavor at their home institution; and the very same researchers may have the hardest time paying the costs of networking under the present structure of research funding, as they may have to bear it entirely on their own. Award levels to investigators must be increased enough to make communications by computer network available to all in the field.

Another aspect of interaction, obvious but vitally important, is face to face communication. This is most feasible on a regular basis where a number of researchers are physically located at the same institution. In this case, organizing themselves into a project group or a larger research center can be a very effective way to promote the needed interdisciplinary communication. As these are very effective structures for interdisciplinary research, we will recommend federal funding for group projects, not necessarily with a single theme or goal, and for at least two major research centers (that is, centers involving at least twenty investigators). Centers have a number of strengths: they allow researchers the flexibility to move quickly into new research problems as they learn new aspects of another discipline, they provide a critical mass of knowledge for projects requiring a particularly diverse mixture of expertise, and they are a national resource facilitating exchange of time-critical information among researchers and between the basic research community and industry and government. Project groups offer some of these advantages, and furthermore are feasible at institutions which do not have a sufficient mass of researchers to constitute a major center.

By supporting such research environments, the federal government can furthermore assist the training of the next generation of researchers in the field—in graduate school, as postdoctoral fellows, and by retraining of senior faculty. This is essential if the field is to fulfill the promise it holds of eventual benefits through application of basic research results to intelligent systems and communications technology.

Estimate of Required Support

In the section of this report titled "Estimate of Needs" (p. 63), we present a rather detailed derivation of an estimate of support required to meet these needs. This includes approximately $5 million in one year for the capital cost of computing equipment, declining to perhaps $1 million a year afterwards. It includes approximately $15 million the first year for continuing expenses of research: the costs of computer maintenance and network com-

munications; salary support for research time, training of graduate students and postdoctoral scientists, and exchange of researchers between institutions; the expenses of workshops and other support of visiting scientists at research centers and larger research groups; other ordinary costs such as travel and materials. Initially up to 150 investigators could be supported, either individually or as a member of a group or center, out of perhaps 500 potentially supportable researchers. As research in the field of Information, Computation and Cognition grows, we anticipate that the annual need for continuing research expenses will increase also, at least doubling over the course of five years, justified by progress of the field.

The estimates just summarized are conservative. The figures mentioned are for direct costs only; they do not include the indirect costs that will be needed in order for institutions to carry out this research.

Need for Additional Support

The support we estimate to be needed for basic research in Information, Computation and Cognition is two to three times the current level of federal funding for relevant research; it exceeds the sum of current federal and private foundation support. Furthermore, existing federal support is mainly directed at the core disciplines rather than at interdisciplinary work in Information, Computation, and Cognition. Clearly a funding shortfall exists now.

The situation will be worse by 1987, however. Funding from the Sloan and System Development Foundations, presently a mainstay of support for the interdisciplinary field, will be ending. Industry cannot be expected to support more than a small part of basic research, though it will provide substantial funding to develop technological applications of the research in question. Some governmental agencies, including several in the Department of Defense, support mission-oriented research, as well as some basic research in the core disciplines; but they have provided very little support for interdisciplinary basic research.

Without a substantial source of new support, this exciting area of research will be in danger of losing its vigor and momentum. We consider it imperative for the federal government to initiate a program of new support for basic research in Information, Computation and Cognition lest the exciting scientific and technological opportunities this field presents be lost.

The interdisciplinary character of the work makes it particularly vulnerable to funding shortages. Existing institutional structures make it easier for investigators to obtain support (material and intellectual) for studies that fall squarely within their home disciplines. Until the ongoing realignment of fields is more firmly established, research that crosses disciplinary

boundaries will need careful nurturing. The potential benefits for science and technology are, as we shall indicate, well worth the cost.

In recommending a new program of interdisciplinary funding, we wish to emphasize again that this cooperative research would be impossible without flourishing core disciplines, and would soon wither without them. Progress in interdisciplinary research remains crucially dependent on continued research support for these disciplines, and since existing funding for the disciplines is relatively modest, it is important to understand that the estimates of need in this document are for *incremental* support.

Technological Need for Basic Research

Developments in Information, Computation and Cognition are affecting the nature of technology, and dramatically changing the theories that illuminate and guide it. It has now become clear that information technology requires information theory, and that although the elements of this theory are found in many existing disciplines, future progress depends crucially on interdisciplinary cooperation. A number of research projects in the "Examples" section of this report show not only the usefulness of the new theories, but the fruitfulness of interactions with technology, which often have suggested new problems and points of view to theoretical workers. What seems to be emerging, and should be fostered by science policy, is a healthy situation in which both technology and basic scientific research benefit from each other.

Every one of the illustrations of research listed in the "Examples" section consists of a theoretical research problem that has manifest applications to the design of systems that perform intelligent tasks, to the facilitation of communication, and to the enhancement of interactions between humans and machines. Some of these examples have led to systems that are commercially available. It is clear enough why this is to be expected.

As computer programs become more intelligent, and computational resources enable us to store vast amounts of information, it becomes increasingly evident that the usefulness of this computational power depends crucially on understanding the specific human needs and thought processes of the user. Just as an encyclopedia is useless unless the information is divided into natural units, and properly indexed, a database is useless unless it is organized in a way that renders accessible the information a user may require. We can judge the complexity of the problem by considering that even in well organized libraries, many users' problems can only be solved through the mediation of an expert reference librarian, who performs a task that in general requires much experience and judgment, not only regarding the structure of the library, but also concerning the intellectual contours of the field, and the needs of human users.

As databases become increasingly large, their usability diminishes proportionally, due to the difficulty of automating this task. Our point is that this problem cannot be addressed without using the techniques of many disciplines, and advancing our understanding not only of information and how to organize it for machine storage, but of the structure of natural language queries for information, of the human knowledge representation of the subject matter, and of the likely goals of the user. Mathematically rigorous theories of the sort that are beginning to develop in the field of Information, Computation and Cognition will be essential to the solution of these problems.

Similar problems emerge in virtually every area of information technology. As programs become increasingly complicated, for example, the need for ways of making them more transportable—more easily adapted to new computing environments and more readily applicable to analogous domains—becomes more urgent. And it is becoming clear that transportability is closely connected to intelligibility, which in turn is related to the way that humans represent algorithms and reason about them. Again, an interdisciplinary approach seems most promising.

As with any basic scientific research, it is impossible to tell in advance which particular ideas and projects will prove to be most valuable for technology, especially in the long run. But it is possible to identify trends, as in this case: both technological needs and the dynamics of the theoretical problems themselves indicate that strongly interdisciplinary approaches are of vital importance.

The large corporations in the information industry demonstrate their expectation that basic research in Information, Computation and Cognition will prove useful by employing researchers to help transfer the results into applied research and product development groups. Three of these companies were represented at the workshop. The section of this report titled "Role of Industrial Research Centers" (p. 83) makes clear the perceived need among industrial researchers for a healthy program of basic scientific research relating to their work.

At present the United States leads the world in the basic research that may be translated into technology within about ten years. It enjoys this lead thanks to past work much of which was supported by private foundations. Current knowledge can be expected to reach the limits of its technological usefulness within this time frame, however. A continuing program of basic research is necessary to develop new knowledge capable of supporting the powerful and intelligent information processing systems that are widely expected for the future. Some other countries, most notably Japan, seem not yet to have understood the need for a sustained program of basic research

on Information, Computation and Cognition, though they are very astute at exploiting the short-term applications of present knowledge. They cannot be expected to remain in the dark for long, however. If the US fails to secure its lead by instituting a strong national program of support for the crucial basic research, it may lose its lead in this critical area of national endeavor. It is vital for the federal government to carry out its role in the partnership of university, industry and government by supporting the necessary basic research.

Conclusion

Recent years have marked the emergence and rapid growth of a highly interdisciplinary area of research in Information, Computation and Cognition involving at least the disciplines of artificial intelligence, computer science, linguistics, logic, philosophy, and psychology. Collaborative research activities in this area, which currently have considerable momentum, offer promise of major advances on basic scientific and foundational questions which are not only of central importance to the core disciplines but will crucially affect the future of technology in the areas of language, information, computation, and communication.

The exciting prospects of the new field are seriously thrown in doubt by the current funding picture. Much of the growth has been supported by two private foundations, the Sloan Foundation and the System Development Foundation, which together have put approximately $46 million into the development of this area since 1977. By 1987, support from those sources will end. New sources must become available if the field is to survive and grow. While National Science Foundation support through present structures has been and will continue to be a significant and steady part of the funding of this research, we believe that a new initiative is needed for the field to have any chance of realizing its potential on a national level.

From the data documented in this report it appears that a beginning program level of $20 million in FY 1987 could provide a substantial base of support. We anticipate that continued growth of the field, scientific progress, and evidence of its value to technological developments would justify doubling that amount over a five-year period, to a rate of $40 million by FY 1992. It is crucial that such support supplement current programs rather than compete with them, since progress in interdisciplinary research depends heavily on the continued strength and vitality of the contributing disciplines.

We see federal involvement as essential also for insuring that this be a truly national effort, not one confined to a small number of privileged institutions. The seed money of the two private foundations went to about a

dozen institutions, in concentrated doses that were appropriate to the start-up phase of such research but would not be a healthy pattern of funding over a longer term.

We have identified a range of kinds of support that we believe are essential for developing and maintaining an appropriate level of scientific work in this area.

- Computer networking facilities to connect researchers across institutions and across disciplines. This is particularly important in the case of such interdisciplinary research, and will require a major effort to implement on a national scale.

  $0.5 million per year
- Computer equipment. The current disparity in computational resources across disciplines and institutions is a major obstacle to the development of an integrated national research community.

  $5 million in the first year, $1 million each succeeding year
- Computer operating expenses.

  $0.75 million per year
- Graduate and postdoctoral support, the latter perhaps in part nationally administered. Essential for the development of the future generation of researchers in this area, who may otherwise be forced by financial exigencies to choose other careers.

  $1.5 million per year for postdoctoral support (steady state)

  $3 million per year for graduate students (steady state)
- A spectrum of research support, including reasonable research funding for at least two major research centers, grants to collaborative group projects and to individual investigators whose research is part of the larger interdisciplinary effort. These are equally high priorities. The level of research support should rise with activity in the field. To facilitate long-range planning and stability of research and training, three- to five-year grants are important.

  $10.475 million per year
- Provision of partial salary support for visiting researchers, perhaps in conjunction with individual and group research grants, to promote intensive collaborative work across a substantial number of institutions.

  $1.875 million per year

The figures given are for direct costs only. The total cost of carrying out the program will be higher by the amount of indirect costs.

From the ready consensus the panel achieved on these recommendations, and from informal consultation with many colleagues, we are fairly confident

that these recommendations would meet with strong approval among the wider research community.

We urge that steps be undertaken as quickly as possible to develop a major new initiative to support research, training, and facilities in the field of Information, Computation and Cognition, at a level that will enable it to fulfill its scientific potential and serve the national need for the future technology that depends on tomorrow's scientific progress.

Without such an initiative this country risks losing its leadership position in a field which is at the threshhold of potential explosive growth internationally as well as nationally. With it, we can expect to see major advances in our understanding of both minds and machines, of language and communication, of the nature of information and its acquisition, transmission, and processing—advances which are urgently needed in this Age of Information.

History of the Report

By the spring of 1985, many researchers had become aware that an increasingly interdisciplinary effort was emerging, focused on language, information, cognition and computation. A sense of excitement about the content and promise of the research was tempered by concern about the likely effects of expected funding patterns on its future. Therefore, the time seemed ripe for a workshop that would take stock of where this field stood, where it was headed, and what its needs were likely to be. To this end, the National Science Foundation and the System Development Foundation sponsored a two-day Workshop on Information and Representation in Washington, D.C., March 30 through April 1, 1985.

The Workshop was chaired by Barbara H. Partee from the University of Massachusetts. Participants included a panel of researchers from universities and from industrial research centers, representatives from private foundations and members of the National Science Foundation staff. The program and list of participants follow at the end of this section.

At the meeting, it became clear that a report of the workshop's deliberations would be most useful if published very quickly. The four individuals who presented reports at the meeting were asked to prepare them for the published proceedings. As for the remainder of this report, committees were organized to draft its various sections, in time to be circulated to the entire panel for comment and revision before publication. The Table of Contents lists the authors of each section; the sections which were so circulated are marked by an asterisk. Overall editorial responsibilities rested with Barbara Partee, Stanley Peters and Richmond Thomason, Peters being responsible for coordinating the editorial process.

The panel members each have their individual training, research experience, and opinions. In approaching our task, we have tried to transcend the limits of our personal histories and present circumstances as much as possible, but we recognize that these must affect our perspective on scientific needs and funding priorities. Another panel with different members might have had different views, just as strongly held. We have tried to guard ourselves against parochialism through frank and searching debate, informal consultation with many colleagues, and by a constant awareness that any recommendations we might make would be subject to the scrutiny of our colleagues in the field.

In general, the areas of consensus are much broader than the areas in dispute. And in many of the latter we feel that the health of the field

would be best served by fostering multiple approaches to problems. As we proceeded in search of meaningful consensus, we have learned much from the staff of the National Science Foundation, from each other, and from colleagues outside the group. But the final product must be taken for what it is: the considered collective judgments of a particular group of scientists, drawn from most of the core disciplines contributing to the study of Information, Computation and Cognition and a number of areas within them, hoping to represent all our hundreds of colleagues, but not presuming to have succeeded.

Schedule of Events

Saturday Evening, March 30: Hotel Lombardy, Conference Room

7:00–11:00 Open House

Sunday Daytime, March 31: State Plaza Hotel, Classroom

10:00 Welcoming Remarks *Barbara Partee*

10:10 Background of the Workshop *Stanley Peters*

10:30 Discussion of the Research Area *All Participants*

- What are the fundamental research questions?
- What theories and methodologies does the research draw on?
- What research projects are currently underway?
- What is the focus and what are the boundaries of the area?
- What consequences can progress in this area of science have for other areas of science and for the development of technology?

12:00 Short break

12:15 Discussion of Research Environment *All Participants*

- What institutional settings are appropriate for this research?
- What community of researchers is involved in the work?

1:00 Level and Profile of Current Funding *All Participants*

2:00 Discussion of Planned Funding Trends other than Federal Support *All Participants*

3:00 Adjourn

Sunday Evening, 31 March: Hotel Lombardy, Conference Room

7:00–10:00 Open House

Monday Daytime, April 1: Joseph Henry Building, National Academy of Science, Room 451

9:00	Summary Presentations Presentations by *John Perry* *Jerry Fodor* *Barbara Partee* *Mitch Marcus*	
11:00	Discussion of Focus and Boundaries	*All Participants*
12:30	Lunch	
1:15	The Current Funding Situation Presentations by *Eric Wanner* *Carl York*	
1:45	Mechanisms for Support for the Research Presentations by *Lila Gleitman* *Richmond Thomason* *Jay Keyser* *Robert Ritchie*	
2:15	Discussion of Research Organization and Funding Structure	*All Participants*
3:00	Divide into working groups to draft sections of the report.	
5:00	Adjourn	

Addresses

Hotel Lombardy 2019 I Street, NW Washington, D.C.	Park Plaza Hotel 2117 E Street, NW Washington, D.C.	Joseph Henry Building, Room 415 National Academy of Science 2122 Pennsylvania Avenue, NW Washington, D.C.

Changes to Program

Please note a number of changes from the published schedule: Participants elected to use Sunday night's open house mainly as a drafting session; on Monday presentations were also given by Jaime Carbonell (an example of problem-oriented research) and by Richmond Thomason (an illustration of how to collect and organize data on current funding); and Jay Keyser presented the report of the group which drafted the afternoon session on Mechanisms for Support for the Research, and chaired the discussion of it.

Participants

Panelists

Stephen Anderson
Department of Linguistics
UCLA

Jaime Carbonell
Department of Computer Science
Carnegie-Mellon University

Herbert Clark
Department of Psychology
Stanford University

David Evans
Department of Philosophy
Carnegie-Mellon University

Jerry Fodor
Department of Linguistics
and Philosophy
Massachusetts Institute
of Technology

Donald Foss
Department of Psychology
University of Texas - Austin

Lila Gleitman
Department of Psychology
University of Pennsylvania

Clark Glymour
Department of Philosophy
Carnegie-Mellon University

Barbara Grosz
Artificial Intelligence Center
SRI International

Aravind Joshi
Department of Computer Science
University of Pennsylvania

Hans Kamp
Department of Philosophy
University of Texas - Austin

Samuel Jay Keyser
Department of Linguistics
and Philosophy
Massachusetts Institute
of Technology

Mitch Marcus
Linguistics and AI
Research Department
AT&T Bell Laboratories

David MacQueen
Computing Science Research Center
AT&T Bell Laboratories

Gary Olson
Department of Psychology
University of Michigan

Barbara Partee
Department of Linguistics and
Department of Philosophy
University of Massachusetts
Amherst

John Perry
Department of Philosophy
Stanford University

Stanley Peters
Department of Linguistics
Stanford University

Stanley Petrick
IBM Thomas J. Watson
Research Laboratory
Yorktown Heights, New York

Robert Ritchie
Computer Science Laboratory
Xerox Palo Alto Research Center

Stanley Rosenschein
Artificial Intelligence Center
SRI International

Dana Scott
Department of Computer Science
Carnegie-Mellon University

Richmond Thomason
Department of Linguistics and
Department of Philosophy
University of Pittsburgh

Thomas Wasow
Department of Linguistics and
Department of Philosophy
Stanford University

Invited Representatives of Private Foundations

Eric Wanner
Cognitive Science Program Director
Sloan Foundation

Carl York
Director of Program Administration
System Development Foundation

National Science Foundation Staff Members

Fred Betz
Industry/University
Cooperative Research
Projects Program

Charles N. Brownstein
Divison of Information
Science and Technology

Paul Chapin
Linguistics Program

Kent K. Curtis
Division of Computer Research

Caroline M. Eastman
Information Science Program

David Kingsbury
Directorate for Biological,
Behavioral, and Social Sciences

Richard T. Louttit
Division of Behavioral and
Neural Sciences

Lawrence Oliver
Special Projects Program
Division of Computer Research

Joseph Young
Memory and Cognitive
Processes Program

Examples of Interdisciplinary Research Projects and Topics

Preface

These examples were collected from the workshop participants by Aravind Joshi and Stanley Rosenschein. They are but a sample of the work going on in Information, Computation and Cognition. We have had the responsibility for the final stage of preparation of this section; given the time constraints, we chose to work on developing the examples on the list, rather than widening the sample, as a limited number of examples fairly well described seemed more likely to convey the excitement and the interdisciplinary nature of the work. We have hopes, however, that this document might be regarded as but the first edition of something that might grow, with time, into a fairer survey of Information, Computation and Cognition, and might perhaps serve as the basis for an ongoing "clearinghouse" that would keep researchers informed of developments throughout the field.

Many hands have contributed to the development of this section; nearly all of the workshop participants provided examples. Jaime Carbonell, Barbara Grosz, Aravind Joshi, Hans Kamp, Jay Keyser, Marcy Macken, Mitch Marcus, David MacQueen, Gary Olson, and Moshe Vardi provided references and other information, often on short notice. Joan Bresnan, Herbert Clark, John Etchemendy, Janet Fodor, Robert Moore, Stanley Rosenschein, Peter Sells, and Susan Stucky performed major writing and rewriting tasks, as did Barbara Partee, Stanley Peters, Richmond Thomason, and Thomas Wasow who also provided much needed direction and encouragement. David Israel did all of the above and more. The responsibility for any remaining inadequacies are our own.

—Elizabeth Macken and John Perry

Introduction

These examples were collected from the participants at the workshop. They are intended to illustrate but by no means exhaust the sort of interdisciplinary research that comprises Information, Computation and Cognition. Moreover, in some cases the pressure of time has made it impossible to collect all the appropriate references and citations. We have somewhat arbitrarily sorted the examples into the following four categories:

(1) The Structure of Representational Systems, Human and Machine
(2) Discourse and Problems of Context

(3) Learning and Reasoning in Humans and Machines
(4) The Nature of Informational Content

Projects listed under heading (1) deal with the representational structures that encode information, both those internal to minds or machines and those used in communication. Projects listed under (2) consider the effect of context on content, particularly in the study of discourse. Projects listed under (3) focus on the acquisition of information by minds or machines, and on the planning and execution of actions by intelligent agents. Projects listed under (4) are concerned primarily with the nature of information; this heading encompasses most of what is more commonly called semantics. Each category cuts across research done in several of the traditional academic disciplines; under each, we list collaborative efforts and indicate some of the questions that must be answered by further research. Table 1, which appears at the end of this section, shows for each collaborative effort, the traditional academic disciplines in which the principal investigators were trained.

Concerning funding, it is important to note that while some funding for related work in individual disciplines has been provided by DoD, NIH, NIMH, and NSF, many of the ideas central to the projects discussed would not have been developed without the interdisciplinary support from the Sloan Foundation and the System Development Foundation (SDF). The joint work among psychologists, linguists, philosophers, and computer scientists encouraged by these foundations has been a principal factor in the current lively state of Information, Computation and Cognition. For example, many of the researchers cited here were not yet in graduate school when the Sloan program began, and have since benefited from Sloan Fellowships and Sloan and SDF supported postdocs. The careers of these researchers would have taken a quite different course without these programs. No current government program fully provides the needed support for such interdisciplinary and group research.

1. The Structure of Representational Systems, Human and Machine

People and machines perceive, learn, reason, and plan. That is, they acquire information from their environment, reorganize it in various ways, and finally bring that information to bear on subsequent interactions with the environment. In this process, the information must be encoded internally, whether in the brain or in silicon. A large part of psychological research is devoted to establishing exactly what information humans must have in order to do what they do, and a large part of research in artificial intelligence (AI) aims to establish what information must be provided a computer or robot if it is to do what we want it to do. In both cases, our understand-

ing of the agent's internal structure is both constrained and illuminated by our understanding of the linguistic structures used to communicate with or program the agent. (See also [Marcus]* for a discussion of structural parallels between natural and computer languages.) This is because the three structures—psychological, computational, and linguistic—are related both causally and through the informational content they share.

1a. The Structure of the Mental Representation System

As discussed in [Fodor] (see also [Partee]), merely establishing what information is internally encoded is not enough. In order to study the properties of the internal processes that manipulate information in perception, reasoning, and planning, we must also know how the information is encoded. However, determining how information is mentally 'packaged' can be extremely difficult, for systems with exactly the same information differently packaged will in many ways exhibit the same externally observable behavior. Such informationally equivalent systems must be characterized and distinguished.

One contribution to the goal of distinguishing informationally equivalent systems is to generate conceptions of different things that might, in principle, serve as representational systems. Philosophers, psychologists, linguists, and computer scientists have explored a range of possibilities, from iconic systems, such as those involved in visual imagery (Shepard and Metzel, 1971; Shepard 1978, 1982; Shepard and Cooper, 1982; Kosslyn, 1980; Marr, 1982; Marr and Poggio, 1976), to highly articulated abstract codes, such as the formulae of intensional logics (Johnson-Laird, 1983; Osherson and Rips, 1984). (For intensional logics, see Montague, 1974.) The current indications, centered around the hypothesis of the modularity of mind (Fodor, 1983), are that different codes may be involved in different psychological faculties.

Another research task is to identify abilities and behaviors sensitive to differences in representational format, and which thus could reveal its nature. For example, people make certain sorts of deductive inferences quickly and accurately, while others seem quite universally to cause considerable difficulties. This suggests to some the use of 'mental models' quite unlike standard logical representations of propositions (Johnson-Laird, 1983). Similarly, the size and character of the stages in language acquisition that take a child from ignorance to full command of the adult language may hold clues to formal properties of the rule systems in which linguistic information is encoded (Fodor and Crain, in progress). See Section 3a.

---

* Names appearing in square brackets refer to one of the workshop presentations; these are reprinted in the last sections of these Proceedings.

1b. Grammars of Natural Language

Much current work in grammatical theory is embedded within frameworks developed not solely for their value as research tools, but as abstract systems representing the human language capability. In particular, the interplay between constraints derived from empirical studies and constraints imposed by considerations of the formal power and computational complexity of grammatical systems has led to a striking convergence among linguists working within several frameworks concerning various features a grammar should have.

These frameworks are usually referred to with acronyms; we provide here a brief glossary:

- Generalized Phrase Structure Grammar (GPSG) provides a single level of syntactic representation produced by a context-free phrase structure grammar and associates with that representation a model-theoretic semantic interpretation; it was developed by Gerald Gazdar, Ewan Klein, Geoff Pullum, Ivan Sag, Carl Pollard, Thomas Wasow, and others at various institutions in the US and abroad, including Stanford University, the University of California at Santa Cruz, Hewlett-Packard Laboratories (H-P), the University of Sussex, and the University of Edinburgh.
- Lexical Functional Grammar (LFG) investigates the role of functional information (primitives such as 'subject', 'object', etc.) in linguistic description expressed at a level of 'functional structure' which is essentially a projection of lexical information; it was developed by Joan Bresnan and Ron Kaplan at the Massachusetts Institute of Technology (MIT), Xerox Palo Alto Research Center (Xerox PARC) and later at the Center for the Study of Language of Information (CSLI).
- Functional Unification Grammar (FG) has had a strong influence on the approaches mentioned above, in that all use the fundamental operation of 'unification' in their implementations; it was developed by Martin Kay at Xerox PARC.
- Categorial Grammar was invented by the Polish logician Ajdukiewicz and developed by Lambek, Curry, Bar-Hillel, and others in the 1950s. It is used in some Montague grammars and makes the fit of syntax to semantics quite rigid by using function-argument application as the method of syntactic combination. Emmon Bach, at the University of Massachusetts, is developing an extended categorial grammar which is richer than the classical categorial grammar but more constrained than Montague's uses. It has been extended to the use of function composition by A. Ades and M. Steedman (1982).

- PATR is a framework within which one can embed, and hence compare, alternative grammatical formalisms, as well as alternative grammars within those formalisms; it was developed by Jane Robinson, Stuart Shieber, Lauri Karttunen, Fernando Pereira, Susan Stucky, and Hans Uszkoreit (Shieber et al., 1984) at SRI International (SRI)/CSLI.
- Government-Binding Theory (GB) provides multiple levels of syntactic representation with constraints among the levels. GB was developed by Noam Chomsky and his colleagues at MIT. In GB, lexical information is projected into domains that use structural information almost exclusively, rather than the feature-value or attribute-value approach to the representation of information which is characteristic, respectively, of GPSG and LFG.

Recent developments in GPSG have led to a revision pioneered by Pollard, known as the Head Grammar (HG) system. This system takes much of the work out of the syntax per se, placing more emphasis on information stored as part of lexical entries. In this way many of the ideas in HG are very similar to those in LFG and also in GB.

Recent developments in LFG, HG, and FG have been directed toward providing an interface to a semantic description using Situation Semantics, thereby allowing a wider range of semantic facts to be accomodated.

Computationally, HG is an extension of GPSG in that it goes beyond context-free power, an extension recent research has shown to be empirically forced. Other research in this area involves the Phrase-Linking Grammars developed by Stanley Peters and Robert Ritchie at the Universities of Texas and Washington, which employ a mild extension of context-free grammars by allowing daughters to have two mother nodes in the syntax, and the Tree-Adjoining Grammars (TAGs) of Joshi of the University of Pennsylvania, which provide a similar extension by allowing pieces of trees generated by a context-free grammar to be combined in novel ways.

As noted above, this apparent embarrassment of riches is only apparent; the different names are labels for theories with essentially the same concerns—descriptive adequacy, computational tractability, and mathematical elegance. Indeed, the first project we mention involves becoming clearer about the similarities and differences.

Comparison of Grammatical Formalisms. Syntactic theories over the past two decades have produced an impressive array of frameworks, formalisms and terminology. Various subsets of these have struck many as in some deep sense empirically equivalent. But there has been no clear basis for comparison, whereby genuine theoretical differences could be separated from artifacts of the particular formalism or notation.

This winter, a number of computer scientists and linguists met regularly at CSLI to compare four syntactic theories, FG, GPSG, HG, and LFG, that are under active development there. Real progress on this "demarcation problem" was made. Shieber was able to show how each of the alternatives could be embedded within some part of PATR, thus demonstrating the striking degree of theoretical convergence between the various frameworks. Kay developed a software package called Basic Linguistic Tools, which permits analyses from all five approaches to be implemented and tested within a neutral computational setting. A dividend of these explorations of the demarcation problem was the development of more efficient parsing algorithms for some of the candidates; one idea, due to Kay, increased the speed of operation of the head-driven parser by a factor of forty.

These discussions have had a dramatic impact on the theorists and the theories as well. They could not have taken place except in a multidisciplinary and multi-institutional setting; in this case linguists and computer scientists from Stanford, Xerox PARC, H-P, and SRI were involved. We might note that an extensive computer network would make this sort of interaction possible even among institutions not in the same geographical area. It is still the rare linguist or philosopher who has access to such networks.

In a way similar to the PATR enterprise, Kroch and Joshi (1985) have demonstrated that GB and other current grammatical frameworks can be instantiated within the TAGs, which, with the demonstration of near-equivalence of TAGs and HGs (discussed immediately below), allows for unified discussion and investigation of grammatical formalisms.

Constrained Grammatical Systems. The array of syntactic theories described above share the constraints imposed by the elimination, in whole or in part, of the transformational component of the grammar. Thanks to the work of Kelly Roach (1984) and K. Vijay Shankar and Aravind Joshi (1985), it has now been established that all known formal properties of tree-adjoining grammars (TAGs) also hold of Pollard's (1984) HGs, and conversely. Further it has been shown that for any TAG an equivalent HG can be constructed. The converse has not been established, but it is conjectured to be true. Given that TAGs and HGs are built on apparently different principles, these results are indeed very surprising.

The study of such systems of grammars has significant implications regarding the computational complexity of the parsers needed in natural language interfaces, and so is being carried out as a joint enterprise by linguists and computer scientists.

1c. Sentence Grammars and Discourse Grammars

Current linguistic practice imposes a sharp division of labor between 'sentence grammar' and 'discourse/text grammar'. This division is justified under the assumption that the sentence defines a scientifically legitimate unit of research. In practice, proponents of sentence grammar have subscribed to two, partially independent points of view: first, that all the linguistic information necessary to account for a given phenomenon is available within a sentence, and second, that only linguistic information in a rather narrow sense is necessary for this. The assumption has been that structural information, or information about functions like *subject*, *object*, etc., is necessary, but that information about notions like *topic* and *focus*, which are crucial to the coherence of discourse, is dispensable.

While these assumptions have not precluded reasonable accounts of certain phenomena, there are also extremely important linguistic phenomena that cannot be handled within the sentential domain, at least not without losing various insights or incurring unnecessary complications in the grammar. In these domains the rigid division of labor has inhibited progress and led to uninsightful or incomplete accounts. More generally, this division has lent credence to highly inadequate models of what the domain of linguistic knowledge consists in. One of the areas in which the fragmented approach has been least successful is in the study of anaphora and ellipsis; another is in the determination of word-order by sentence-external discourse factors. In other areas the approach has lead to 'benign neglect' of those phenomena that do not fit within the preestablished subdivisions. For example, since a great number of agreement phenomena can be handled in a sentence-internal way, agreement is generally looked upon as a problem of sentence grammar. But some agreement phenomena require a larger perspective, and hence have not been studied extensively.

Anaphora and Discourse Representation Theory. Recent work on anaphora has shown that a joint treatment of intra- and intersentential anaphora yields simplified conditions on possible co-reference. The work of Hans Kamp (1984), and the closely related work of Irene Heim (1982), has brought a significant advance in the treatment of nominal anaphora, utilizing a level of representation in terms of Discourse Representation Structures (DRSs). Kamp's work has generated new approaches to some long-standing puzzles in anaphora, notably those involved in the so-called 'donkey-sentences'; in addition, it points toward a unification in the treatment of pronouns, in that anaphora between sentences is treated in exactly the same way as anaphora within a sentence.

Notable about Kamp's and Heim's work, which was motivated by traditional logical and linguistic considerations, is its appeal to the construc-

tion of local representations that are dynamically updated during discourse. Such an approach is directly relevant to psychological and computational models of discourse processes. (See Grosz and Sidner, in progress) The subject offers many opportunities for fruitful interdisciplinary cooperation.

Discourse and Grammatical Interactions. Recent research within Lexical-Functional Grammar shows that sentence-internal grammatical processes, such as word order and verbal morphology, also interact systematically with discourse functions in ways that can be precisely characterized. Recent work by Bresnan and Mchombo (in progress) on Chichewa (a Bantu language spoken in East Central Africa) suggests principles for syntactically identifying *topic* and *focus* across languages, and hence could provide a useful tool for empirical investigations of the integration of syntactic structure and discourse structure.

Bresnan, Halvorsen, and Maling (in progress) argue that there are clear interactions between the grammatical and the discourse dimensions of anaphoric binding systems. For example, Icelandic reflexive pronouns can be used to refer to an individual in the context whose speech, thoughts, or point of view are reflected in an indirect discourse context. But these pronouns must simultaneously meet the condition that their antecedents be grammatical subjects, even across sentences. This suggests that some sentence-internal grammatical information is preserved in discourse structures.

LFG, as a very explicit and highly modular theory, provides a useful framework from which to study the interaction between discourse and sentence phenomena. Moreover, the general architecture of the framework allows experimentation with different modes of interaction between different components. Linguistic models up to now, LFG included, have displayed a marked preference for the serial approach. However, there is no need for the components of grammars built on unification to interact in a serial rather than a more 'parallel' fashion. The different subcomponents can constrain the output without being in linear order; one such model incorporating the LFG components is proposed in Jens Erik Fenstad et al. (1985).

1d. Problems of Lexical Entry

All speakers of a language must know the words of their language and these words must be represented mentally. Researchers in the Lexicon Project (at the Center for Cognitive Science, MIT) are seeking to learn what constitutes knowledge of a lexical item and how this knowledge is best represented, and to develop a computationally based model of the lexicon corresponding to their theoretical work. They expect to produce a working model of a computer based dictionary, one whose entries are theoretically sophisticated

and which is equipped with sufficient dictionary handling programs to make the lexicon useful, not only as a theoretical tool but as a practical one as well.

A number of languages of quite different structures are being investigated, including English, Berber, Winnebago, and Warlbiri. Researchers are focusing primarily on syntactic and semantic properties of lexical items and are looking particularly at verbs. As yet, there exists no dictionary of Winnebago and a suitable data base is presently being compiled by Josie White Eagle. Research papers presenting various theories of the lexical entry are in various stages of progress; for example, B. Levin and M. Rappaport (1985) and K. Hale and S. J. Keyser (in progress).

1e. From Natural Language to Programming Languages

If we accept the hypothesis that brains are computers, then natural languages must be recognized as our most powerful "programming languages." We can, after all, get a human being to perform a complex task—say, mowing the lawn—by uttering a few words. A similar instruction for a computer would take pages of code. Even if we are doubtful about the computational hypothesis, natural language is clearly an extraordinarily powerful communicative and instructional device, and knowledge of it will prove useful in developing and improving computer languages.

One payoff of understanding the increased complexities of natural language would be the development of programming languages, or in a more limited case, database query languages, far more concise than current languages. While the pervasive ambiguity of natural language argues against the use of unconstrained natural language for specifying algorithms, certain features such as quantification and superlatives are relatively easy in natural language, though difficult and unwieldy in formal query languages.

The problem of developing natural languages as programming or database query languages is being studied in a number of contexts. Warren and Pereira (1982) compared Chat-80, a fragment of English, to Quel, one of the most concise and user-friendly languages for database systems. Ballard's (Bierman and Ballard, 1980; Ballard and Tinkham, 1984; Ballard, 1982) NLC, a formal English subset for specifying algorithms or matrices, was quickly taught (within 45 minutes) to undergraduates who had just completed a PL/C programming course. It was then used by the students to implement algorithms more quickly and accurately than they had implemented the same algorithms in PL/C. [Marcus] gives a similar example from the LADDER System. In contrast, SQL, a database query language, requires the user to know the SQL syntax, the db structure (names of db relations, keys of relations, etc.), and specific codes. In general, the natural languages provide a more compact representation of queries.

1f. The Implications of Different System Architectures for Cognitive Processing

Work on natural language processing has thus far proceeded on the assumption that algorithms will be implemented on sequential machines. There is now considerable interest in looking at algorithms with a view toward implementing them on parallel processing machines, as well as algorithms explicitly designed for implementation on these new architectures. For example, Joshi (Joshi and Palis, in progress) has been investigating parallel processing for TAGs and the associated semantic computations. Other investigators have begun to explore parallel, connectionist models (McClelland and Rumelhart, in progress; Hinton and Anderson, 1981; Feldman and Ballard, 1982). There are clear examples where models based on these different architectures yield divergent accounts of a particular phenomenon.

Focused interdisciplinary research will be needed to take full advantage of the information to be gained from the new system architecture. The work promises eventually to yield very efficient implementations for natural language interfaces. As a by-product, it will provide new models for processing that may turn out to be both linguistically and psychologically significant.

2. Discourse and Problems of Context

As emphasized in [Partee] and [Perry], problems of discourse and context have been one of the chief spurs to recognition of the inherently interdisciplinary (and extremely complex) problems that confront any attempt to build natural language systems. Even a cursory examination of the way languages, whether natural or artificial, are actually used reveals clearly that extended sequences of utterances are the norm, not single statements.

Moreover, while discourses exhibit internal structure much like sentences, the constituent structure is not determined by the linear sequence of utterances. It is common both for two contiguous utterances to be members of different subconstituents of the discourse, and for two noncontiguous utterances to be members of the same subconstituents (e.g., see Grosz, 1981; Reichman, 1978). This structure plays an important role in the processing of utterances in the discourse. The structure in turn depends on several factors, including the syntactic and semantic properties of the individual utterances and the mental states of the participants. Each participant brings to the discourse a set of beliefs, goals, intentions, and other mental attitudes all of which influence how utterances are produced and how they are understood.

Our interactions with one another constitute an ongoing process, not a single event. Discourses are not static objects, but rather the result of agents acting in a particular way to affect changes in the world. A compu-

tational theory of discourse must provide both an account of how utterances affect the discourse situation and an account of how the discourse situation affects the generation and production of utterances. The mental states of the participants and the preceding discourse are crucial ingredients in such accounts.

By regarding a discourse as a basic unit of analysis, we affect a change in the way one looks at language in general. Not only are new theoretical (and computational) constructs needed to handle specific mechanisms of discourse, but the treatment of syntactic and semantic phenomena is altered by being considered as part of this larger endeavor. For example, at the semantic level, representations of the meaning of noun phrases must encode information sufficient for interpreting subsequent references (for a discussion of what this entails, see Weber, 1978). Again, utterances that are equivalent syntactically and/or semantically may well not be from the perspective of the discourse.

The problems just mentioned in explaining and interpreting discourse among human participants occur in parallel form in our communication with machines. The typical program requires many lines of code; in processing each line may interact with neighboring or distant lines of code, the effects of the interactions depend on the internal states of the machine, and the results affect the processing or the next line of code. See [Marcus] for a description of some similarities between notions of discourse and the block structure constructs of ALGOL-like programming languages.

Many of the observations just made have appeared in the literature of philosophy, sociolinguistics, and other disciplines, but it is only recently that researchers in computational linguistics and AI have begun to develop computational theories capable of accounting for them. As demonstrated in the XCALIBUR project at Carnegie-Mellon University (CMU) (Carbonell, Boggs, Mauldin, and Anick, 1983; Carbonell and Hayes, 1983), such studies are crucial for practical applications as well as for their theoretical import. The examples below illustrate some specific problems in understanding discourse phenomena; in some cases computational approaches to solutions are in sight, but not in others.

2a. Discourse and Reference.

In the early 1970s, Wallace Chafe (1974) argued that pronouns could be used in ordinary conversation or narratives only when the objects they referred to were in "consciousness" (as he put it). This topic, the interactions among the referential structure of discourse, grammar, and the psychological states of speaker and hearer has been a continuing source of interdisciplinary

investigation, where insights about discourse structure are modeled computationally and tied conceptually to work in semantics of natural language.

In her contribution to an important volume on discourse understanding Barbara Grosz (1978/ in Joshi, Weber, and Sag, 1981) showed in some detail how, in task oriented discourse, "focusing," as she called it, reflects the structure of the task and how the choice of referring expression affects and is affected by focus. She introduced the notion of a focus space, and showed how to represent focus spaces. This approach has been developed in papers by Sidner (1981), Reichman (1982), and others. More recently, Grosz, Joshi, and Weinstein (1983) have developed the computational notion of centering to deal with a number of intricate problems in pronominalization. These ideas have recently been applied in the LFG framework (see "Discourse and Grammatical Interactions" in Sec. 1c) by Megumi Kameyama (1985) to the zero (i.e., silent) pronouns in Japanese, and they are being tested in psychological studies of the production and understanding of pronouns by Clark and his students.

The volume in which Grosz's paper was published grew out of a Sloan sponsored conference at the University of Pennsylvania. A number of other papers in this volume stimulated investigations of related discourse phenomena. Bonnie Weber (1978) took as central the notion of the model of a situation that the speaker was trying to direct the listener to synthesize, and related such models to semantic analyses of various puzzling cases by linguists and philosophers (Karttunen, 1976; Partee, 1972; Hintikka and Carlson, 1977). Raymond Perrault and Phil Cohen (1978/ in Joshi, Weber, and Sag, 1981) approached the problem of inaccurate reference from a speech act perspective, and Clark and Marshall (1981) studied the "mutual" or common knowledge constraint on reference.

These researchers have assumed a number of different constraints and a range of ideas about what is most central in the interplay of discourse structure and reference. Grammatical, content, discourse, and processing constraints have played different roles in the different approaches. Many of the resulting theories are compatible, but while a number of different interdisciplinary groups now have suggestions for how they might be merged, a great deal of work remains to flesh these out. For example, for each type of expression, such as singular quantified noun phrases or plurals, researchers must be concerned with the wide range of uses and contexts, the interaction of the different constraints, and the determination of which constraints get brought to bear when.

All of these approaches to discourse point to the general phenomenon of the effect of contextual facts of various sorts—the preceding discourse, the shared goals of the participants, their models of each others' under-

standing, the physical setting, and so forth—in actual communication. The implication that the interactions of all of these factors with language have to be understood before we can understand our ability to use language to communicate and achieve other goals has been a major stimulus to much of the interdisciplinary work described elsewhere in this section.

2b. Indexicality

Philosophers, linguists, and psychologists have long been concerned with the nature and representation of indexicals (e.g., 'I' and 'you') and demonstratives (e.g., 'that' and 'this'). David Kaplan (1977) has argued persuasively that indexicals refer to things in the world not via descriptions of those things, but via context. John Perry (1977, 1979) and David Lewis (1979) have argued that indexicality is not a superficial phenomenon, but reflects the intrinsically perspectival nature of psychological states. George Miller (1982) used these arguments to suggest that the current psychological representations of what people say are fundamentally inadequate. According to Miller, these representations need to characterize what he called the 'appearance' of things—the speaker's perspective on those things. Geoffrey Nunberg (1979) developed a detailed analysis of indirect uses of these indexicals, as when someone points at an empty chair and says, "That man is returning in five minutes." Clark and Robert Schreuder (Clark, Schreuder, and Buttrick, 1983) elaborated on Nunberg's analysis to handle apparently indeterminate uses of these pronouns, as when one points at TWO men and says, "That man is fat." The whole issue of indexicality and propositional attitudes has been examined by Robert Stalnaker, most recently in his book, *Inquiry*, which develops a treatment of the attitudes within the possible worlds framework (see Sec. 4a, and [Perry]).

2c. The Problem of Common Knowledge

As discussed in [Marcus], communication clearly requires knowledge of one another's intentions, practices, and the like, but only recently have researchers recognized the full import and scope of this requirement (Grice, 1957; Lewis, 1969; Shiffer, 1972; Clark and Marshall, 1978/in Joshi, Weber, and Sag, 1981). Suppose A refers to an object in talking to B. For understanding to take place, B must know which object A refers to. But this is not enough: A must know that B knows this. And B must know that A knows this. And so on. Thus an infinite set of conditions seems buried under an innocent act of successful reference. Some researchers have seen such an infinite set or an infinitely complex condition as essential to the analysis of common knowledge; Clark and Marshall suggest instead that the central fact is that a single situation is both what is known and the knowing of it.

The problems associated with the notion of "common knowledge" are being studied by a number of investigators. For example, Cohen (1978) has suggested a simple representation for common knowledge usable for computational purposes. More recently, Halpern and Moses (1985) have demonstrated how the same notions can be used in analyzing network or communication protocols in distributed information processing systems. Jon Barwise (1985c) has developed a theory in which common knowledge, instead of having an infinitely complex representation in terms of internal connections, is capable of simple unitary representation of the kind Clark and Marshall suggest.

2d. Illocutionary Acts

The analysis of what are technically called 'illocutionary acts' (that is, assertions, questions, apologies, commands, and the like) came originally from John Austin (1962), John Searle (1969), and, in a different framework, from Paul Grice (1957). These philosophers conceived of language as part of purposeful action. The linguistic expression of illocutionary acts became a serious topic of investigation in the 1970s (with linguists Sadock (1974), Gazdar (1979), and others). A central question in this research was, how so-called indirect illocutionary acts, like "Do you know what time it is?" are interpreted in context.

Clark (1979) and Gibbs (1981), psychologists, have investigated several theories of how indirect requests are understood in ordinary contexts. In AI, there had been a tradition of work on planning in intelligent systems, and it was Cohen, Perrault, and Allen (Cohen and Perrault, 1979; Perrault and Allen, 1980) who first attempted to integrate this work with the philosophical notions of purposeful action and to account for indirect illocutionary acts. Their work was followed by Cohen and Levesque's (1980) on the logic behind the speakers' plans of action, by Sidner and Israel's (1981) on how listeners recognize speakers' plans, and by Doug Appelt's (1982) on how speakers reason about their listeners' mental states (to plan referring expressions). One type of utterance examined by Allen and Perrault (1980) were sentence fragments like "The 3:15 to Montreal?" as said to a railway ticket seller. This has spawned psychological research into the mental processes by which these fragments are understood in context.

2e. Cooperative Principles

Grice (1975) has proposed that much of what we understand in conversation is not to be found in what the speaker literally says, but in what he or she implies (or "implicates," to use Grice's term). In conversation we take for granted that everyone is being cooperative; in particular, we assume that the speaker is being as informative as necessary, truthful, relevant, and clear.

So when someone says, for example, "There were no philosophy students in Linguistics 10 last year," we assume that there were at least some students in the course and that it was taught, though neither inference is required. We draw these conclusions on the assumption that the speaker was being as informative as possible.

The cooperative principle has a wide range of consequences that are only now being studied. However, Grice did not develop these ideas in the direction of providing detailed general mechanisms or algorithms for inferring implicature or, more generally, for inferring a speaker's intended meaning from his or her utterances. The work on planning described in Section 3g. provides promising first steps toward such mechanisms and algorithms. The crucial idea is related to Grice's; the focus, however, moves from Grice's very general maxims of cooperativeness to the dynamic construction of detailed models of the beliefs, plans, and goals of the conversational participants. This area provides an example in which theories designed to meet computational needs can usefully influence such disciplines as linguistics and philosophy.

We mention the following as examples of the many research topics bearing on implicature.

Preventing False Inferences. In cooperative man-machine interaction, it is necessary that a system (S) truthfully and informatively respond to user's (U) question. It is not, however, sufficient. In particular, if S has reason to believe that its planned response might lead U to draw an inference that S knows to be false, then S must block it by modifying or adding to that response. Because information is communicated implicitly in discourse as well as explicitly, a user may well be misled by what the system does not say. Because U expects S to block certain conclusions that S does not know to hold or knows to be false, if S does not block them, U will assume the conclusions are justified. Thus, not saying something, i.e., not blocking unwarranted conclusions that the user might draw, can be misleading. The issue is related to default reasoning and nonmonotonic reasoning (see Sec. 3d), as these inferences can be called conversational defaults.

An extension of this research is the question of how we constrain S's behavior in this respect. We do not want S to prevent every possible false inference the user could draw. How should one localize this behavior? Research in this area requires close cooperation between AI researchers, linguists, and philosophers. These and other problems of preventing false inference are being explored by Joshi, Weber, and their students (Joshi, 1982).

Deducing the Right Questions. By studying many transcripts of people engaged in task-related discourse, it has been found that people do not always ask for precisely the information they need because they do not

know how to. Moreover, they know they don't know how to, and hence expect a cooperative, informed respondent to be able to deduce from what seems to be the goal of the query (communicative goal) their intended, unstated (and perhaps by them unstatable) domain goal. The system has to be given expertise about the domain and about the common user goals within it to be able to deduce the user's domain goals, given U's incomplete and perhaps inappropriate query. Martha Pollack (1984) has been pursuing these problems drawing heavily on Goldman's work on logic of action, the planning work in AI (see Sec. 3g), Joshi (Nadathur and Joshi, 1983) and the related work of Bonnie Weber (1983).

Scalar Implicatures. One type of implicature is so-called "scalar" implicature. Suppose person A is asked, "How many brothers do you have?" and he answers, "Two." Logically, A could give this answer truthfully even if he had three, five, or one hundred brothers. Yet, because we assume he is being cooperative, we assume he has given us the highest point on the scale he truthfully could, and we infer that he has two brothers and no more. This notion of scale and implicature extends to other attributes and other implicatures—e.g., "Policemen in Britain must be six feet tall" implies they must be *at least* six feet tall. These implicatures have been studied by Hirschberg (1985), Hirschberg and Ward (in progress), Fauconier, and Horn. This example of scalar implicatures raises the general problem of specifying a useful taxonomy of implicatures; some work in this direction has been done.

3. Learning and Reasoning in Humans and Machines

Over the last thirty years, the conjecture that all human intelligence could in principle be simulated by a digital computer, and the bolder hypothesis that minds are in fact computers, have been the focus of a great deal of controversy and fruitful work in philosophy, computer science, linguistics, and psychology. In a seminal article, Alan Turing devoted considerable attention to the possibility of the development of machines that could learn from experience, as humans do (see [Partee]). He concluded there was no problem in principle, but that this would be one of the most difficult aspects of human intelligence to simulate in a machine. These conjectures have resulted in a tremendous flow of ideas back and forth among the disciplines. Here we give only a brief sketch of the areas of accomplishment and current interest in uncovering learning and reasoning abilities in humans and instilling them in machines.

3a. The Learnability of Grammars

As Chomsky has pointed out, one of the most remarkable facts about natural

language is the ease and speed with which children learn their first language.

In the early 1970s, K. Wexler and P. Culicover (1980), and H. Hamburger developed a mathematical model of language learning based on a version of transformational grammar. On the basis of certain assumptions about the nature of the input data to the learner and the structure of the acquired grammar, they were able to prove that a linguistically plausible class of grammars was learnable in principle, from simple data (in a rigorously defined sense of "simple").

The Wexler, Culicover, and Hamburger research explored the consequences for learnability of specific assumptions about the form of grammars and the input to the learner. Osherson, Weinstein, and Stob (1983) have explored the same issue from a somewhat more abstract perspective, as have Joshi and Levy (1978), using a different approach. In particular, they have proved a series of theorems illuminating the relationship between learnability in principle and the initial conditions on grammar form and learner input. This work provides one clear framework within which to formulate and test learnability hypotheses.

There have been interactions between work in formal learning theory and empirical studies of language acquisition in children. For example, Lila Gleitman (Landau and Gleitman, 1985) has studied language acquisition in children with severely restricted input capacity (i.e., children who are blind or deaf).

3b. Machine Induction

The problem of inducing a grammar or language from a finite sample is a particular instance of the more general problem of inducing functions from argument value pairs, and more generally still, of generalizations from instances.

Programs that form theories about bodies of data (e.g., Shapiro's Model Inference System) typically must perform computationally expensive tests of hypotheses. It is essential to efficient performance that computation time not be wasted in generating and testing hypotheses that are entailed by hypotheses already established or in generating and testing hypotheses that entail hypotheses that are already rejected. This problem was recognized some years ago by Gordon Plotkin (1972), but efficient algorithms for solving the problem have not been developed. To avoid the computational problems, programs such as Shapiro's restrict the language they use and the hypothesis space search, thus reducing the power of the program. This problem is partly logical and partly computational, and of considerable long-range significance.

3c. From Learning to Reasoning

Since the problem addressed in formal learning theory and machine induction is discovering the appropriate generalization from instances, any implementation must rely on heuristic algorithms. More generally, research in AI has brought sharply into focus the centrality of heuristics in reasoning. This brings to light the need for a treatment of such reasoning comparable in scope, rigor and utility to deductive logic or probability theory (see also [Partee]).

As classically conceived, rules of deductive or statistical reasoning are not domain specific. Much of the work in AI, however, strongly suggests that heuristic reasoning is domain specific (see, for example, the work of Roger Schank (1975, 1982) and Robert Abelson (Schank and Abelson, 1977)). What do domain specific rules of inference look like? Why would a cognitive system rely on such domain specific rather than general rules? How do general rules come out of domain specific rules? These and related questions account in part for the large role that philosophers—whose interests are sometimes thought to be too abstract and other-worldly to be relevant to even the most theoretical of sciences—are playing in the development of Information, Computation and Cognition. For example, the problem of updating probabilities given uncertain evidence, addressed by the philosopher Richard Jeffrey, among others, is relevant to the development of machines that use probabilistic reasoning. And, the BACON programs developed at CMU were certainly facilitated by Herbert Simon's background in philosophy of science.

Another traditional concern of philosophers is reasoning by analogy. Its importance has been recognized by philosophers from Aristotle to Hume and Mill; and Polya has studied its role in mathematical problem solving. Recently, within the AI paradigm, attempts have been made to model processes of analogical reasoning by Jaime Carbonell (1983, 1985), Dedre Gentner (undated, 1983), Patrick Winston (1979), and others. This is an area the investigation of which requires collaborative effort by logicians, psychologists, and AI researchers.

The flow of ideas here is not simply from philosophy and logic to computer science, however. The computational perspective provides new insights on old problems. For example, Clark Glymour has argued that the design and analysis of programs for learning, confirmation, and reasoning generate useful constraints on and metrics for theories in the philosophy of science.

3d. Nonmonotonic Reasoning

Another area in which developments in AI recall traditional philosophical

ideas and shed technical light on them is *nonmonotonic reasoning*. Out of these trends, exciting new theories are emerging that present alternatives to probabilistic models of learning and inductive reasoning, and that also give a prominent role to logical techniques.

Nonmonotonicity refers to the idea that a conclusion that is warranted on certain evidence may well not be warranted on learning more. My conclusions on hearing a fire alarm in a hotel may be overridden when I see a sign announcing a test.

One source of nonmonotonicity can be traced to rules of reasoning that admit exceptions. The idea that even scientific rules can be defeasible goes back to Aristotle, and the notion of a *prima facie truth* is common in the philosophical literature. However, the first formally adequate model of nonmonotonic reasoning to appear, and the one that presently prevails in philosophical circles (and certainly, in statistical and economic circles) is probabilistic. In this context, the nonmonotonicity of reasoning is represented by the fact that the conditional probability of B on A can be very low, even though the probability of A is very high. Of more general significance in this regard, however, has been the work of Jeffrey and others on the updating of probabilities under the pressure of new, typically uncertain, evidence.

Despite the intellectual attractiveness and mathematical tractability of this approach, it has not been seen as useful by many computer scientists, largely because it does not seem to motivate reasoning algorithms. It is very difficult, for instance, to see how to model a probability function computationally. However, expert systems not only employ defeasible rules, but combinations of these rules so complex that these implementations soon created a great need for theories in terms of which they could be understood.

In constructing such theories, computer scientists turned to the formalisms of symbolic logic. One leading approach, due to Jon Doyle and Drew McDermott(McDermott and Doyle,1978; Doyle,1979a; Doyle,1979b; McDermott, 1980; Doyle, 1982), is an application of modal logical techniques. They introduce a modal operator governed by a nonmonotonic rule of inference and define validity in terms of fixpoints of this operator. The definition renders inferential relations non-effective in an important sense: there will, in general, be no effective test of whether a sequence is a proof. This approach has been refined and developed in various ways; see especially the analysis by Robert Moore (1983). And in his later work on truth maintenance and reasoned assumptions, Doyle (1983) has generalized the ideas in numerous directions, and provided philosophical depth.

A few philosophers and logicians have begun to contribute to this work. Moore's research was influenced by unpublished work of Stalnaker's (1980); and more recently, Dov Gabbay (1982) has explored adaptations in intu-

itionistic logic of the modal approach to nonmonotonicity. The modal approach is now a well-established research program, with a rich logical texture and many documented applications in computer science.

An alternative approach of John McCarthy's (1980a, 1980b) is based on the idea of circumscribing the range of predicates in a standard first-order theory by adding postulates formulated in higher-order logic. These postulates can be conceived of as akin to the second-order induction axiom for arithmetic. This idea applies with particular immediacy to presuppositions that enter into problem-solving activity; but it can be generalized to apply to many, and perhaps all forms of nonmonotonic reasoning. This approach is younger than the modal one, but it lends itself well to logical treatment and has suggested some interesting problems in mathematical logic the solutions to which have produced some illuminating theorems. Again, it presents a very fruitful area in which technical work in logic can interact productively with computer science. (See, for instance, Lifshitz, 1984.)

The computer scientist Ray Reiter (1978, 1980) had earlier developed a related approach in which the language remains that of standard first-order logic. Rather than add to such theories explicit closure axioms, however, Reiter introduces a nonmonotonic rule-schema, instances of which he calls "default" rules and analyzes the behavior of standard first order theories under the operation of default inference. Recently, he has been studying the relations between default theories and circumscription. (Etherington, Mercer, and Reiter, 1984.)

How to draw a principled and useful line between deductive and inductive logic is a long-standing issue in philosophical logic. In contemporary logic, this distinction is usually marked by the difference between deductive and statistical relations. However, some technical work has provided other ways of approaching the difference: C. Alchourron's, P. Gärdenfors' and D. Makinson's work on theory change is a notable example (Gärdenfors, 1981; Alchourron and Makinson, 1982; Alchourron, Gärdenfors, and Makinson, 1984; Gärdenfors, 1984). Building in part on this work, and in part on modal theories of the conditional (in which nonmonotonicity appears in the logical properties of the antecedent) Richmond Thomason (Thomason, 1983; Glymour and Thomason, 1984) has recently suggested a theory of default reasoning in which default rules are expressed by conditionals, and their effect is felt through nondeductive rules of theory perturbation. Reflecting on the problem of revising a database in the light of new information, some computer scientists have also thought of using theories based on conditional logic. Matthew Ginsberg's work (1984) is a good example.

These approaches to nonmonotonic reasoning comprise a growing subfield of research, in which computer scientists, mathematical logicians, and

philosophers can interact productively. It suggests numerous research projects, ranging from small-scale technical problems to large, strategic issues. An example of the latter is the integration of default reasoning with planning and action. On the classical Bayesian approach, this is accomplished through the notion of expected utility; on the nonmonotonic reasoning approach, we need some idea of acting on a belief in accordance with a certain desire or plan (for instance Doyle, 1982). Yet another problem is how to monitor nonmonotonic reasoning to maintain a measure of its credibility in the face of risk. Such a measure is certainly needed for acting on defeasible conclusions in the face of risk; the question is, to what extent it approximates a probability function. (See also Sec. 3g for connected developments.)

Interestingly enough, the idea that deductive logic could be applied to inductive reasoning came from computer scientists, and logicians have only recently begun to be influenced by it. As this influence grows, we are likely to see a theory of decision making under uncertainty taking shape that challenges the well-entrenched probabilistic models. The long-term effects of these developments are difficult to predict, but may well extend to all the sciences that presently deal with human and computational decision making.

3e. Reflective Systems

Reflection about one's own knowledge and ignorance is an important and interesting locus of nonmonotonicity. For instance, people often reason that if they don't know something of a certain kind—e.g., that they have a brother—then it must not be so. This inference is nonmonotonic because if we add the premise that one was adopted and never told of one's family, for example, the conclusion can no longer be drawn. For another example, a system might answer "I don't know" on the basis of knowledge of the limits of its own proof system, when asked about the theoremhood of a sentence which, in abstraction from those limitations, it can prove (see also [Partee]). The ability to reflect underlies much of the subtlety with which humans deal with the world. It is essential to mastering new skills and it enables us to cope with the limited nature of our knowledge of the world. Hence another direction for research is the design of systems with a measure of self-knowledge, which, in turn, requires a better understanding of the nature and limits of self-knowledge in humans and other biological species.

Brian Smith (1984) has developed a version of LISP called 3-LISP that embodies "reflection" in a very direct way. In 3-LISP one can write programs which instruct the computer to "step back" and reflect on the processes it was just running or the environment it was running them in, and allows the program to alter either in powerful ways.

Smith is using ideas from 3-LISP in developing a new language MANTIQ that will include a declarative as well as procedural component. A computer language satisfying the goals of MANTIQ would significantly advance the state of the art in AI, cognitive psychology, computer science, and even the philosophy of mind.

3f. Logics of Knowledge and Belief

In the early 1960s, the philosophical logician Jaakko Hintikka (1962) developed formal logics for knowledge and belief. Recent work in this area has been done by Stalnaker (1984), Thomason (1980), and Wolfgang Lenzen (1978). In the mid 1970s, AI researchers including McCarthy (1977) and Moore (1980, 1985) recognized that it was important for intelligent systems to be able to reason about the knowledge and belief of other agents and began integrating Hintikka's ideas into AI work. Moore's work combined Hintikka's treatment of knowledge with a theory of action based on earlier work of McCarthy and Hayes (1969), so that a system could reason about what it would have to know in order to act effectively, as well as about which actions it could take to acquire knowledge. It was then recognized that this work was specially relevant to problems of natural-language generation, and Appelt (1982) developed a generation system that used Moore's formalism to plan utterances, extending previous work of Cohen, Perrault, and Allen (1982) that applied AI techniques for planning and problem solving to speech act planning. More recently, AI researchers Kurt Konolige (1984, 1985), Levesque (1984), and Israel (in progress) have developed formal theories of knowledge and belief that permit weaker assumptions about the reasoning abilities of agents than the approaches of Hintikka, McCarthy, and Moore do. Also theoretical computer scientists including Fagin, Halpern, and Vardi (1984) and Halpern and Moses (1985) have been developing logics of knowledge to analyze the flow of information in distributed computer systems.

3g. Practical Reasoning, Planning, and Action

Philosophers since Aristotle have made a distinction between practical and theoretical reasoning. Theoretical reasoning leads from belief to belief; practical reasoning from beliefs and goals or desires to action.

The most influential modern approach to practical reasoning is decision theory (Jeffrey, 1984). Actions are assigned expected utilities on the basis of the probability of the relevant beliefs and the utilities of outcomes contingent on those beliefs. Decision theory has played a large role in economic theory, and the mathematics of decision theory is fairly well understood. The underlying notions of belief and action raise thorny conceptual problems, however. Some of these have to do with the structure of action and of

the objects of belief. Goldman's (1970) analysis of action has been influential in this regard. He analyzes action in terms of basic acts which generate non-basic action in various causal or conventional contexts. For example, moving one's left arm in a certain way, a basic action, may conventionally generate the action of signalling for a left turn; if done too vigorously, it may causally generate dislocating one's arm and ruining one's summer vacation. Goldman's theory appreciates the circumstantial nature of action; that what one does is determined not only by basic actions but the circumstances in which they occur. Recent work in the philosophy of language and epistemology has emphasized the circumstantial nature of belief and desire. This relates to the phenomenon of indexicality considered in Section 2b.

The connection between the circumstantial nature of action and the attitudes can be seen in actions closely associated with one's cognitive attitudes. For example, by saying "I am sitting," *A* says that he is sitting; *B*, uttering the same words, says something different, that he is sitting. The same context relativity applies, it seems, to the underlying cognitive states; this is a point made by the "twin earth examples" of Hilary Putnam and others (Perry, 1979). Moore's (1985) theory of knowledge and action, and recent work by Perry (1985), suggest ways of accomodating the interconnections between the context relativity of the attitudes and action.

A second set of problems have to do with the role of intention in practical reasoning. Michael Bratman (1981, 1983, 1984, 1985) has argued that intentions cannot be reduced to desires or some combination of desires and belief, but play an important and irreducible role in practical reasoning, as a way of providing stable commitments that allow for the development of plans that play sort of a default role in guiding one's action; one does not constantly update one's plans in the light of changing desires and beliefs; such updating is itself a goal-directed activity.

In the AI context, the phenomenon of practical reasoning has been studied in the framework of planning and program generation. One can think of planning as generating a program on the basis of information and goals, which then guides the action of machine or human. There is a rich literature on this approach, beginning with works on planning as deduction by McCarthy (1958), Hayes (McCarthy and Hayes, 1969), Green (1969), and others. Planning is conceived of as proving theorems about actions that are sufficient or optimal for achieving certain goals in the light of certain information and in such a way that these theorems provide a program for action. This approach has been used in the design of robots, such as Shakey-I at SRI (Fikes, Hart, and Nilsson, 1972).

A seminar held at CSLI in the winter of 1983 provides a good illustration of the utility of interdisciplinary cooperation between philosophers

and computer scientists. The seminar compared philosophical approaches to practical reasoning and the planning literature in AI. The work of Bratman and Perry mentioned above, the work of Moore discussed in Section 3f, and the work by Rosenschein and Pereira described in Section 4c was all influenced by this seminar.

3h. Connected Developments in Logic and Theoretical Computer Science

A number of logicians, going back to the 1950s, noticed that there was a strong analogy between formulae and types on the one hand, and proofs and programs (expressed as terms in the lambda calculus) on the other. This analogy was exploited in De Bruijn's Automath project (1983) to formalize and verify proofs in classical mathematics, and it is also the basis for Per Martin-Löf's Intuitionistic Type Theory (ITT) (1983), an attempt to formalize constructive, or intuitionistic, logic.

These systems have inspired several applications in computer science. For instance, the PRL system of Robert Constable and J. Bates (1972) at Cornell University is a more or less direct implementation of a logic related to ITT. It performs proof-checking and semi-automatic theorem proving in this logic, and in addition is able to extract reasonable programs from proofs of formulae specifying the intended behavior. ITT has also been used as a program specification language by a group of researchers at the University of Gothenburg (Nordstöm and Petersson, 1983). Another less direct application is that type-theoretic ideas from ITT and Automath have proved an appropriate basis for language constructs supporting modularization and "programming in the large" in languages such as Russell (Donahue and Demers, 1980), ML, and Pebble (Burstall and Lampson, 1984).

Temporal logics are playing a role in developments in computer science, which in turn are stimulating the development of new temporal logics. They are proving useful as hardware specification languages since change of state is an essential feature of hardware operation. Temporal logic also finds application in the specification and development of asynchronous programs. The challenges in these areas are leading in turn to the development of new temporal logics geared to the specific problems of computer science. So far, to our knowledge, there has been no substantial interpenetration of this research with the equally intensive research involving temporal logic and natural language semantics (van Benthem, 1983; Dowty, 1979; Kamp, 1970), but the intensive and productive activity now going on in these areas makes them ripe for even more strongly interdisciplinary development.

Among the most important developments in theoretical computer science has been the work on denotational semantics for programming lan-

guages, work which stems from Dana Scott's (1973, 1982) provision of models of various sorts for the full untyped lambda calculus and other related formalisms for combinatory algebras. The crucial problem to be solved here lay precisely in the untyped nature of these formalisms, in particular in the lack of a prohibition against self-application for functions. Research on untyped systems, in turn, has led more recently to research on flexibly typed formalisms—the so-called polymorphic or variably-typed systems. In this guise it has already begun to influence work on the semantics of natural languages, especially work on nominalizations. (See Turner, 1983; see also Sec. 4a.)

Other new logics which have been or are getting developed out of concerns arising in computer science include: Joseph Goguen and José Meseguer's equational logic (1981, 1983), algorithmic logic, dynamic logic, and Hoare's logic of programs.

4. The Nature of Informational Content

As emphasized in [Fodor] and [Perry], the development of a mathematically rigorous and philosophically cogent theory of information is a central problem for Information, Computation and Cognition. In a sense, theories of information have long played a role in the study of language, for semantics assigns models, possible worlds, or other candidates for informational content to expressions or utterances on the basis of lexical and syntactic properties. Adequate theories of cognitive and computational structure should allow semantical investigations of these systems, and hence contribute to a general account of information flow across humans and machines.

4a. Possible World Semantics and the Syntax of Natural Language

A particularly successful approach to the study of the semantics of natural language was begun by the philosopher of language Richard Montague (1974). Due to the efforts of Partee (1972, 1976) and Thomason, linguists, along with philosophers, have subsequently developed a body of literature that treats not only semantical issues (such as the intension/extension distinction and the semantics of belief) within the possible worlds semantics that Montague espoused, but which has proven to be a fertile ground for investigating the relation between syntax, semantics, and pragmatics. For instance, Montague's work spawned a development of the Gricean program by Karttunen and Peters (1979), in which an account of conventional implicature was provided in a Montague-like fragment.

Much of the work in the Montague tradition on the interaction of syntax and semantics centers on questions like these: Can semantical issues impinge on syntactic analysis, and which sort of analysis, a semantic or a syntac-

tic one, provides the best understanding of particular linguistic phenomena? For example, some have thought that the semantic correlate of an infinitival should be a proposition. This requires an rather abstract syntactic analysis, since on the surface the materials for a proposition aren't available. Montague showed how to give a property correlate for infinitivals, allowing a more intuitive syntactic treatment. In recent work, Chierchia (1984) has proposed innovations to Montague's analysis to treat a wide class of nominalizations.

The effect that Montague semantics has had on syntactic theory has been a broad one, touching all the major syntactic frameworks. Categorial grammar, which Montague himself used, has been expanded in a number of ways (by Bach (1979), for example); it has also figured into the work of Ades and Steedman (1982), who use it as the basis for their psychological model of human grammatical competence. Montague semantics has also been paired with grammars in the transformational paradigm, starting with Partee (1973). Recent work by Mats Rooth (1985) exploits the more recent version of that theory, Government-Binding theory, which he uses together with a Montague semantics to provide an analysis of focus words like 'even' in English. Other major syntactic frameworks, notably LFG and GPSG, have also been paired with a Montague semantics (see Halvorsen (1983) and Gazdar et al. (1984)). This has been a fruitful tradition and one that promises to continue to promote understanding not only of natural language semantics but of syntax and pragmatics, and the interactions among these various components.

4b. Situation Theory

Situation theory is a theory of information content that can be seen either as an alternative to, or development of, possible worlds semantics. Situations differ from possible worlds in being partial, and throughout the theory the power of partial functions is used. Some of the ideas were developed by Barwise and Perry in *Situations and Attitudes* (1983). The development in that book was in some respects sketchy, and the application to natural language was not very well developed. The interaction with computer scientists and linguists in the CSLI setting has led to these further developments:

- An improved theory of conditionals (Barwise, 1985b).
- A development of the theory as an alternative to set theory rather than as a definitional extension of set theory (Barwise and Perry, 1984 and Barwise, 1985a). One impetus for this development was the efforts by Barwise and computer scientists at Xerox PARC and SRI to develop a situation semantics for programming languages. The inherently richer structure of situations was inhibited for these

purposes, in its set-theoretical version.

- An attempt to develop the beginnings of a theory of attunement to information as a relation between a psychology (formally conceived), an environment, and an interpretation of psychological states as cognitive attitudes (Perry, 1985).
- A new approach to the use/mention distinction (Stucky, 1985).
- Explorations in the organization of data-bases by situations (Goguen, in progress).
- Development of a semantics for Lexical Functional Grammar (see Sec. 2) within situation semantics (Fenstad, Halvorsen, Langholm, van Benthem, in progress).

4c. Situated Automata Theory

A theory of situated automata is being developed by Rosenschein and Pereira (1985). The work grew out of interactions between computer scientists, philosophers, and logicians and represents an attempt to develop a qualitative information theory for machines that interact with their environments. The approach combines elements of automata theory, the logics of knowledge of action (see Sec. 3f), and intuitions developed in situation semantics, especially with respect to indexicality of information. Situated automata theory differs radically from traditional theories as a basis for attribution of semantic content (see [Perry]), in that it allows attributions of propositional content in the absence of a denotation relation defined on elements of an internal "language of thought." Following Fred Dretske (1982), the ascription of content is based on systematic correlations of state between the automaton and the embedding environment. By connecting this idea to precise models of computation, the theory gives a mathematical characterization of how a machine can possess and transform informational content.

Aside from its philosophical interest, the theory also suggests possible new directions in AI design methodology. The traditional AI approach to representation and reasoning views a program's "knowledge" as being embodied in symbolic expressions that function very much like formulae in a logic in that they (a) derive their content from an "interpretation" assigned (at least implicitly) by the designer, and (b) are manipulated "proof-theoretically" by the program during the course of its operation. Situated automata theory suggests that the semantic content of states and processes can be rigorously characterized without having to view the machines as performing inferences on symbolic data. This has direct consequences for the development of efficient AI programs, and the idea is currently being applied to the design of intelligent robots.

4d. Semantically Coherent Programming Languages and Portable Software

As explained in [Marcus], improved theories of content have led to the development of semantically rationalized programming languages (e.g., Smith and des Rivières, 1984). Such languages support constructs that allow access to the inner (virtual) structure of the interpreter itself. These constructs then allow the quick implementation of debuggers that move transparently to implementations on any other machine at all, because the debugger itself is based upon the virtual semantic structure alone.

The point that semantic coherence promotes portability is not restricted to debugging devices. To the extent that the structure and meaning of programs can be specified semantically, portability across environments will be enhanced.

Conclusion

The reader should bear in mind that the foregoing is not an exhaustive survey of work in Information, Computation and Cognition, but merely a sampling not only incomplete, but inevitably oriented towards the interests of the workshop participants and those who prepared the present version. Information, Computation and Cognition is an alive and exciting field, attracting large numbers of graduate students whose work will doubtless make this list dated in a short time.

The projects listed in this section have developed in a number of contexts: in departments of linguistics, philosophy, psychology, and computer science in universities around the nation, in research institutes and industrial laboratories, and in the centers established by Sloan and SDF.

The work so far has benefited greatly from the available means of communication provided by existing computer networks and centers, and by support provided individual researchers by NSF as well as the foundations mentioned. At the same time, it has been inhibited by the relative isolation of many researchers who want to participate, and the relative scarcity of funding for group and interdisciplinary projects—scarcity that will intensify as the initiatives of Sloan and SDF come to an end. As the field develops, communication between different research sites will become even more important, and funds to train graduate students and postdocs, so they can contribute to the cross-fertilization and development of the interdisciplinary research, will be crucial.

Table 1

Home Disciplines of the Principal Investigators Mentioned in the Interdisciplinary Project Descriptions

1. The Structure of Representational Systems, Human and Machine

	AI	CS	LING	LOGIC	PHIL	PSYCH
a. The Structure of the Mental Representation System	√		√		√	√
b. Grammars of Natural Language	√	√	√			√
c. Sentence Grammars and Discourse Grammars			√		√	√
d. Problems of Lexical Entry		√	√			√
e. From Natural Language to Programming Languages	√	√	√		√	
f. The Implications of Different System Architectures for Cognitive Processing	√	√				√

2. Discourse and Problems of Context

	AI	CS	LING	LOGIC	PHIL	PSYCH
a. Discourse and Reference	√	√	√	√	√	√
b. Indexicality			√	√	√	√
c. The Problem of Common Knowledge	√	√		√	√	√
d. Illocutionary Acts	√		√		√	√
e. Cooperative Principles	√		√		√	

3. Learning and Reasoning in Humans and Machines

	AI	CS	LING	LOGIC	PHIL	PSYCH
a. The Learnability of Grammars		√	√	√	√	√
b. Machine Induction	√	√		√		
c. From Learning to Reasoning	√			√	√	√
d. Nonmonotonic Reasoning	√	√		√	√	
e. Reflective Systems	√	√			√	
f. Logics of Knowledge and Belief	√			√	√	
g. Practical Reasoning, Planning, and Action	√	√		√	√	
h. Connected developments in Logic and Theoretical Computer Science		√	√	√	√	

4. The Nature of Informational Context

	AI	CS	LING	LOGIC	PHIL	PSYCH
a. Possible World Semantics and the Syntax of Natural Language			√	√	√	√
b. Situation Theory		√	√	√	√	
c. Situated Automata Theory	√	√			√	
d. Semantically Coherent Programming Languages and Portable Software	√	√				

BEST COPY AVAILABLE

References

Ades, A., and Steedman, M. 1982. On the Order of Words. *Linguistics and Philosophy* 4:517–558.

Alchourron, C., and Makinson, D. 1982. On the Logic of Theory Change: Contraction Functions and Their Associated Revision Functions. *Theoria* 68:14–37.

Alchourron, C., Gärdenfors, P., and Makinson, D. 1984. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. To appear in *The Journal of Symbolic Logic*.

Appelt, D. E. 1982. Planning Natural-Language Utterances to Satisfy Multiple Goals. Artificial Intelligence Center Technical Note 259, SRI International, Menlo Park, Calif.

Austin, J. 1962. *How to do Things with Words*. Cambridge, England: Cambridge University Press.

Bach, E. 1979. Control in Montague Grammar. *Linguistic Inquiry* 10:515–532.

Ballard, B. 1982. A "Domain Class" Approach to Transportable Natural Language Processing. *Cognition and Brain Theory* 5:269–287.

Ballard, B., and Biermann, A. 1979. Programming and Natural Language: NLC as a Prototype. *Proceedings of the ACM National Conference*. Detroit, Michigan. 228–237.

Ballard, B., and Tinkham, N. 1984. A Phrase-Structured Grammatical Framework for Transportable Natural Language Processing. *Computational Linguistics* 10:81-96.

Barwise, J., and Perry, J. 1983. *Situations and Attitudes*. MIT Press.

Barwise, J., and Perry, J. 1984. Shifting Situations and Shaken Attitudes. Report No. CSLI-84-13.

Barwise, J. 1985a. Lectures on Situation Semantics. Winter Quarter, 1984, Stanford University, to appear in S. Stucky and J. Barwise (Eds), CSLI Lecture Notes.

Barwise, J. 1985b. The Situation in Logic-II: Conditionals and Conditional Information. Report No. CSLI-85-21.

Barwise, J. 1985c. Lecture on Common Knowledge. Spring Quarter, Stanford University, Department of Mathematics.

Biermann, A., and Ballard, B. 1980. Toward Natural Language Computation. *American Journal of Computational Linguistics* 6:71–86.

Bratman, M. (in progress) Intentions, Plans and Practical Rationality. Department of Philosophy, Stanford University.

Bratman, M. 1981. Intention and Means-End Reasoning. *The Philosophical Review* 90:252–265.

Bratman, M. 1983. Taking Plans Seriously. *Social Theory and Practice* 9:271–287.

Bratman, M. 1984. Two Faces of Intention. *The Philosophical Review* 93:375–405.

Bratman, M. 1985. Davidson's Theory of Intention. In Vermazen and Hintikka (Eds.), *Essays on Davidson*. Oxford: Oxford University Press.

Bresnan, J., Halvorsen, P.-K., and Maling, J. (in progress) Invariants of Anaphoric Binding Systems. Xerox PARC, Palo Alto, Calif.

Bresnan, J., and Mchombo, S. (in progress) Agreement and Pronominal Incorporation in Chichewa. Department of Linguistics, Stanford University.

Bresnan, J. 1982. Control and Complementation. In J. W. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press.

Burstall, R., and Lampson, B. 1984. A Kernel Language for Abstract Data Types and Modules. In G. Kahn, D. B. MacQueen, G. Plotkin (Eds.), *Semantics and Data Types*. Lecture notes in CS, 173:1–50. Springer-Verlag.

Carbonell, J. G. 1983. Learning by Analogy. In Michalski, Carbonell, and Mitchell (Eds.), *Machine Learning I* (Chapter 5). Tioga Press.

Carbonell, J. G. 1985. Derivational Analogy, A Theory of Reconstructive Problem Solving and Knowledge Acquisition. In Michalski, Carbonell, and Mitchell (Eds.), *Machine Learning II* (Chap 14). Tioga Press and Kaufmann/Morgan.

Carbonell, J., Boggs, W., Mauldin, M., and Anick, P. 1983. XCALIBUR Project Report 1. Technical Report, Carnegie-Mellon University, Department of Computer Science.

Carbonell, J., and Hayes, P. 1983. Robust Multi-Strategy Parsing to Restricted-Domain Natural Language Interfaces. Carnegie-Mellon University, Department of Computer Science, unpublished manuscript.

Chafe, W. 1974. Language and Consciousness. *Language* 50:111–133.

Chierchia, G. 1984. Anaphoric Properties of Infinitives and Gerunds. *Proceedings of the West Coast Conference on Formal Linguistics.* University of California, Santa Cruz, March 16–18, 1984. M. Cobler, S. MacKaye, M. T. Westcoat (Eds.), Stanford Linguistics Assoc. Vol. 3.

Clark, H. H. and Ravn, K. (in progress). Stanford University, Department of Pyschology.

Clark, H. H., and Clark, E. V. 1977. *Psychology and Language.* Harcourt Brace Jovanovich.

Clark, H. H., and Marshall, C. R. 1978. Definite Reference and Mutual Knowledge. Presented at *Conference on Elements of Discourse Understanding*, University of Pennsylvania. (In Joshi, Weber, and Sag, 1981, 10–63.)

Clark, H. H., Schreuder, R., and Buttrick, S. 1983. Common Ground and the Understanding of Demonstrative Reference. *Journal of Verbal Learning and Verbal Behaviour* 22:245–258.

Clark, H. H. 1979. Responding to Indirect Speech Acts. *Cognitive Psychology* 11:430–477.

Cohen, P. R., and Levesque, H. 1980. Speech Acts and the Recognition of Shared Plans. *Proceedings of the Third Biennial Conference of the Canadian Society for Computational Studies of Intelligence.* Victoria, British Columbia, May 1980. 263–271.

Cohen, P. R., and Perrault, C. R. 1979. Elements of a Plan-based Theory of Speech Acts. *Cognitive Science* 3:177–212.

Cohen, P. R., Perrault, C. R., and Allen, J. F. 1982. Beyond Question Answering. In W. Lehnert and M. Ringle (Eds.), *Strategies for Natural Language Processing.* Englewood Cliffs: Prentice Hall.

Cohen, P. R. 1978. On Knowing What to Say: Planning Speech Acts. University of Toronto, Department of Computer Science, Technical Report No. 118, doctoral dissertation.

Constable, R., and Bates, J. 1972. Proofs as Programs. *ACM TOPLAS* 7:137–158.

Davis, M. 1980. The Mathematics of Non-monotonic Reasoning. *Artificial Intelligence* 13:73–80.

De Bruijn, N. G. 1983. A Survey of the Project AUTOMATH. In J. P. Seldin and J. R. Hinley (Eds.), *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus, and Formalisms.* 579–606.

Donahue, J., and Demmers, A. 1980. Data types, Parameters, and Type Checking. *Seventh ACM POPL Proceedings.* Las Vegas, Nevada. 12–23.

Dowty, D. R. 1979. *Word Meaning and Montague Grammar.* Dordrecht, Holland.

Doyle, J. 1979a. A Glimpse of Truth Maintenance. *Proceedings Sixth IJCAI,* 232–237.

Doyle, J. 1979b. A Truth Maintenance System. *Artificial Intelligence* 12:231–272.

Doyle, J. 1982. Some Theories of Reasoned Assumptions: An Essay in Rational Psychology. Carnegie-Mellon University.

Doyle, J. 1983. The Ins and Outs of Reason Maintenance. *Proceedings Eighth IJCAI,* 349–351.

Dretske, F. I. 1982. *Knowledge and the Flow of Information.* Cambridge, Mass.: MIT Press.

Etherington, D., Mercer, R., and Reiter, R. 1984. On the Adequacy of Predicate Circumscription for Closed-World Reasoning. *Technical Report 84-5.* University of British Columbia. Department of Computer Science.

Fagin, R., and Vardi, M. 1985. An Internal Semantics for Modal Logic. To appear in *Proceedings 17th ACM Symposium on Theory of Computing.* Providence, May, 1985.

Fagin, R., Halpern, J., and Vardi, M. 1984. A Model-Theoretic Analysis of Knowledge: Preliminary Report. In *Proceedings of the 25th Symposium on Foundations of Computer Science.* West Palm Beach, Florida.

Feldman, J. A., and Ballard. 1982. Connectionist Models and Their Properties. *Cognitive Science* 6:205–254.

Fenstad, J. E., Halvorsen, P. K., Langholm, T., and van Benthem, J. (in progress) Equations, Schemata, and Situations: A Framework for Situation Semantics. University of Oslo.

Fikes, R. E., Hart, P. J., and Nilsson, N. J. 1972. Learning and Executing Generalized Robot Plans. *Artificial Intelligence* 3:251–288.

Fodor and Crain. (in progress) *Formal Theory of Innate Linguistic Knowledge.* Bradford Books.

Fodor, J. A. 1981. *Representations* Cambridge, Mass.: MIT Press.

Fodor, J. A. 1983. *Modularity of Mind.* Cambridge, Mass.: MIT Press.

Gabbay, D. M. 1982. Intuitionistic Basis for Non-Monotonic Logic, *Lecture Notes in Computer Science.* Volume 139, Apringer.

Gärdenfors, P. 1981. An Epistemic Approach to Conditionals. *American Philosophical Quarterly* 18:203–211.

Gärdenfors, P. 1984. Epistemic Importance and the Logic of Theory Change. To appear in P. Weingartner and G. Dorn (Eds.), *Foundations of Logic and Linguistics.*

Gazdar, G., Klein, E., Pullum, G., and Sag, I. 1984. *Generalized Phrase Structure Grammar.* Oxford: Blackwell Publishing.

Gazdar, G. 1979. *Pragmatics: Implicature, Presupposition, and Logical Form.* New York: Academic Press.

Gentner, D. The Structure of Analogical Models in Science. BBN Technical Report 4451. Bolt, Beranek, and Newman, Cambridge, Mass.

Gentner, D., and Stevens, A. (Eds.). (in press as of 1983), *Mental Models.* Hillsdale, N.J.:Erlbaum.

Gibbs, R. W. Jr. 1981. Your Wish is My Command: Convention and Context in Interpreting Indirect Requests. *Journal of Verbal Learning and Verbal Behavior* 20:431–444.

Ginsberg, M. 1984. Non-Monotonic Reasoning Using Dempster's Rule. *Proceedings of the National Conference on Artificial Intelligence.* Austin, Texas, August, 1984.

Gleitman, L., and Wanner, E. 1982. Language Acquisition: State of the Art. In E. Wanner and L. Gleitman (Eds.), *Language Acquisition*, Cambridge, England: Cambridge University Press.

Glymour, C., and Thomason, R. 1984. Default Reasoning and the Logic of Theory Perturbations. Nonmonotonic Reasoning Workshop AAAI. October 12–17: New Paltz, N.Y.

Goguen, J. A., and Meseguer, J. 1981. Completeness of Many-Sorted Equational Logic. *SIGPLAN Notices* 16:24-32.

Goguen, J. A., and Mesequer, J. 1983. Logic of Subsorts and Polymorphism. Lecture at Seminar on Types, Carnegie-Mellon University, June, 1983. (Paper in preparation).

Goldman. 1970. *Theory of Human Action.* Prentice Hall.

Green, C. 1969. Application of Theorem Proving to Problem Solving. *Proceedings of the International Joint Conference on Artificial Intelligence*. Los Altos: William Kaufman, Inc. 219–239.

Grice, H. P. 1957. Meaning. *Philosophical Review* 66:377–388.

Grice, H. P. 1975. Logic and Conversation. In D. Davidson (Ed.), *Logic of Grammar*. Dickenson. (Originally presented as one of *The William James Lectures*.)

Grosz, B., Joshi, A. K., and Weinstein, S. 1983. Providing a Unified Account of Definite Noun Phrases in Discourse. Technical Note 292, SRI International.

Grosz, B., and Sidner, C. (in progress) Discourse Structure and the Proper Treatment of Interuptions. SRI International, Menlo Park.

Grosz, B. 1977. The Representation and Use of Focus in a System for Understanding Dialogs. SRI Report No. 151, Menlo Park, Calif.

Grosz, B. 1978. Focusing and Description in Natural Language Dialogs. Presented at *Conference on Elements of Discourse Understanding*, University of Pennsylvania. (In Joshi, Weber, and Sag, 1981, 84–105.)

Hale, K., and Keyser, S. J. (in progress) Some Transitivity Alternations in English. Lexicon Project Working Papers. Center for Cognitive Science, MIT, in preparation.

Halpern, J. Y., and Moses, Y. 1985. Knowledge and Common Knowledge in a Distributed Environment. IBM Research Report RJ 4421 (47909) IBM, San Jose, Calif.

Halvorsen, K. 1983. Semantics for Lexical Functional Grammars. *Linguistic Inquiry* 14:4.

Heim, I. 1982. The Semantics of Definite and Indefinite Noun Phrases. University of Massachusetts, Amherst, Doctoral dissertation.

Hintikka, J. 1962. *Knowledge and Belief*. Ithaca, N.Y.: Cornell University Press.

Hintikka, J. 1971. Semantics for Propositional Attitudes. In L. Linsky (Ed.), *Reference and Modality*. London: Oxford University Press. 145–167.

Hintikka, J., and Carlson, L. 1977. Pronouns of Laziness in Game Theoretical Semantics. *Theoretical Linguistics*. 4(1/2):1–30.

Hinton, G. E., and Anderson, J. A. (Eds.). 1981. *Parallel Models of Associative Memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Hirschberg, J. 1985. Generalized Scalar Implicatures and Their Application to Natural Language Interfaces. University of Pennsylvania, Doctoral dissertation.

Hirschberg, J., and Ward, G. 1985. Semantico–Pragmatic Analysis of Fall-Rise Intonation. To appear in *Language* in June, 1985.

Israel, D. 1980. What's Wrong with Non-monotonic Logic? In *Proceedings of the First Annual National Conference on Artificial Intelligence.* Stanford University, Stanford, Calif. 99–101.

Jeffrey, R. 1984. *The Logic of Decision.* Chicago: University of Chicago Press.

Johnson-Laird, P. N. 1983. *Mental Models.* Cambridge, Mass.

Joshi, A. 1982. Mutual Beliefs in Question and Answer Systems. In N. Smith (Ed.), *Mutual Knowledge.* Addison Wesley.

Joshi, A., and Levy, L. S. 1978. Skeletal Automata. *Information and Control Journal.* December.

Joshi, A., and Levy, L. 1982. Phrase Structure Trees Bear More Fruit Than You Would Have Thought. *American Journal of Computational Linguistics* 8:1-11.

Joshi, A., and Palis, M. (in progress) Applications of Iterative Arrays for Parallel Processing of Grammatical Systems. University of Pennsylvania, Moore School, Department of Computer and Informational Sciences.

Joshi, A., Weber, B., and Sag, I. 1981. *Elements of Discourse Understanding.* Cambridge, England: Cambridge University Press.

Kameyama, M. (in progress) Zero Anaphora: The Case of Japanese. Doctoral Dissertation, Department of Linguistics, Stanford University.

Kamp, H. 1968. Tense Logic and the Theory of Linear Order. University of California, Los Angeles. (Doctoral Disseratation.)

Kamp, H. 1979. Events, Instants, and Temporal Reference. In R. Bäuerle, U. Egli, and A. von Stechow (Eds.), *Semantics from Different Points of View.* Berlin: Springer-Verlad. 376–417.

Kamp, H. 1980. Some Remarks on the Logic of Change, Part I. In C. Rohrer (Ed.), *Time, Tense and Quantifiers.* Tübingen, West Germany: Max Niemeyer Verlag. 135–179.

Kamp, H. 1981. Evenements, Representations Discursives et Reference Temporelle. *Language* 64: December.

Kamp, H. 1984. A Theory of Truth and Semantic Representation. In J. Groenendijk, T. Janssen and M. Stokhof (Eds.), *Truth, Interpretation and Information*. Dordrecht: Foris, 1–41.

Kaplan, D. 1977. Demonstratives. University of California, Los Angeles, unpublished draft #2.

Kaplan, J. 1983. Cooperative Responses from a Portable Natural Language Data Base Query System. Volume of M. Brady and R. C. Berwick (Eds.), *Computational Models of Discourse*. Cambridge, Mass.: MIT Press.

Kaplan, R., and Bresnan, J. 1982. Lexical Functional Grammar: A Formal System for Grammatical Representation. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press, 173–281.

Karttunen, L. 1976. Discourse Referents. In J. McCauley (Ed.), *Syntax and Semantics*. Vol. 7. New York: Academic Press.

Karttunen, L., and Peters, S. 1979. Conventional Implicature. In C.-K. Oh and D. Dinneen (Eds.), *Syntax and Semantics*. 11: Presupposition. New York: Academic Press, 1–56.

Kay, M. 1984. Basic linguistic tools: Software Package. Xerox PARC, Palo Alto, Calif.

Kay, M. 1979. Functional Grammar. *Proceedings of the Berkeley Linguistics Society*. Vol. 5.

Konolige, K. 1984. A Deduction Model of Belief and its Logics. SRI International Artificial Intelligence Center Technical Report-1984-326. Menlo Park, Calif.

Konolige, K. (in progress) Belief and Incompleteness. In J. R. Hobbs and R. C. Moore (Eds.), *Formal Theories of the Commonsense World*. Norwood, New Jersey: Ablex Publishing Corp.

Kosslyn, S. M. 1980. *Image and Mind*. Cambridge, Mass.: Harvard University Press.

Kripke, S. A. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

Kroch, A., and Joshi, A. K. 1985. Linguistic Relevance of Tree Adjoining Grammars. Technical Report No. MS-CIS-18-85. University of Pennsylvania, Department of Computer and Information Science.

Landau, B. and Gleitman, L. 1985. *Language and Experience: Evidence from the Blind Child*. Cambridge, Mass.: Harvard University Press.

Lenat, D. B. 1984. Computer Software for Intelligent Systems. *Scientific American* 251:204–213.

Lenzen, W. 1978. Recent Work in Epistemic Logic. *Acta Philosophica Fennica* 30:1–219.

Levesque, H. 1984. A Logic of Implicit and Explicit Belief. *Proceedings of the National Conference on Artificial Intelligence.* University of Texas. Austin, Texas: August 6–10:198–202.

Levin, B. (Ed.). 1985. *Lexical Semantics in Review.* Lexicon Project Working Papers, Number 1, Center for Cognitive Science, MIT, Cambridge, Mass.

Levin, B. and Rappaport, M. 1985. The Formation of Adjectival Passives. Lexicon Project Working Papers #2, Center for Cognitive Science, MIT, Cambridge, Mass.

Lewis, D. K. 1969. *Convention.* Cambridge, Mass.: Harvard University Press.

Lewis, D. K. 1979. Attitudes De Dicto and De Se. *Philosophical Review* 88:513–535.

Lifschitz, V. 1984. Some Results on Circumscription. Nonmonotonic Reasoning Workshop. AAAI. October 17–19. New Paltz, N.Y.

Lyons, J. 1977. *Semantics* Vols. 1 & 2. Cambridge, England: Cambridge University Press.

Marr, D., and Poggio, T. 1976. Cooperative Computation of Stereo Disparity. *Science* 194:283–287.

Marr, D. 1982. *Vision.* San Francisco.

Martin-Löf, P. 1983. Constructive Mathematics and Computer Programming. *Sixth International Congress for Logic, Methodology, and Philosophy of Science.* Hannover, West Germany, August 22–29, 1979: North Holland.

McCarthy, J. and Hayes P. J. 1969. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer, and D. Michie (Eds.), *Machine Intelligence 4.* Edinburgh, Scotland: Edinburgh University Press, 463–502.

McCarthy, J. 1977. Epistemological Problems of Artificial Intelligence. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence.* Massachusetts Institute of Technology. Cambridge, Mass.: Aug. 22–25:1038–1044.

McCarthy, J. 1958. Programs with Common Sense. In *Proceedings of the Symposium on the Mechanization of Thought Processes, National Physical Laboratory* I:77–84. (Reprinted in M. L. Minsky (Ed.). 1968. *Semantic Information Processing.* Cambridge, Mass.: MIT Press, 403–409.

McCarthy, J. 1980a. Addendum: Circumscription and Other Non-monotonic Formalisms. *Artificial Intelligence* 13:171–172.

McCarthy, J. 1980b. Circumscription–A Form of Nonmonotonic Reasoning. *Artificial Intelligence* 13:27–39.

McClelland, J. L., and Rumelhart, D. E. (Eds.). (in progress) Parallel Distributed Processing: Explorations in the Micro-Structure of Cognition. Hillsdale, N.J.: Lawrence Erlbaum Associates.

McDermott, D., and Doyle, J. 1978. Non-Monotonic Logic I. MIT AI Memo 486. (Also published in 1980. *Artificial Intelligence* 13:41–72.)

McDermott, D. 1980. Nonmonotonic Logic II. Research Report 174, Yale University, Department of Computer Science, New Haven, Conn.

Miller, G. A. 1982. Some Problems in the Theory of Demonstrative Reference. In R. J. Jarvella and W. Klein (Eds.), *Speech, Place, and Action: Studies in Deixis and Related Topics.* New York: Wiley.

Minsky, M. 1974. A Framework for Representing Knowledge. MIT Artificial Intelligence Laboratory, AIM-306. MIT, Cambridge, Mass.

Montague, R. 1974. *Formal Philosophy: Selected Papers of Richard Montague.* R. Thomason (Ed.). New Haven and London: Yale University Press.

Moore, R. C. 1983. Semantical Considerations on Non-monotonic Logic. *Proceedings of the Eighth Conference on Artificial Intelligence.* (Also published in 1985. *Artificial Intelligence* 25:75–94.)

Moore, R. C. 1980. Reasoning About Knowledge and Action. Artificial Intelligence Center Technical Note 191. Menlo Park, Calif: SRI International.

Moore, R. C. (in progress) A Formal Theory of Knowledge and Action. In J. R. Hobbs and R. C. Moore (Eds.), *Formal Theories of the Commonsense World.* Norwood, New Jersey: Ablex Publishing Corp.

Nadathur, G., and Joshi, A. 1983. Mutual Belief Conversational Systems: There Role in Refering Expressions. *Proceedings of the International Joint Conference on Artificial Intelligence.* Karlsruhe, West Germany, August.

Non-monotonic Reasoning Workshop, (October 17–19, 1984) Sponsored by the American Association for Artificial Intelligence, Mohonk Mountain House, New Paltz, New York.

Nordstöm, B., Petersson, K. 1983. Types and Specifications. *Proceedings of IFIP*. R. E. A. Mason (Ed.). Elsevier.

Nunberg, G. 1979. The Nonuniqueness of Semantic Solutions. Polysemy. *Linguistics and Philosophy* 3:143–184.

Osherson, D., and Stob, M., and Weinstein, S. 1983. Environments for Learning. *Information and Control.*

Osherson, D., and Weinstein, S. 1983. Formal Learning Theory. In M. Gazzaniga and G. A. Miller (Eds.), *The Handbook of Cognitive Neurology.* Plenum Press.

Partee, B. (Ed.). 1976. *Montague Grammar.* Academic Press.

Partee, B. 1973. Some Transformational Extensions of Montague Grammar. *Journal of Philosophical Logic* 2:509–534. (Also in Partee, 1976, 51–76)

Partee, B. 1972. Opacity, Coreference, and Pronouns. In Harman and Davidson (Eds.), *Semantics of Natural Language.* Dordrecht and Boston: Reidel.

Perrault, C. R., and Cohen, P. R. 1978. It's for Your Own Good: A Note in Inaccurate Reference. Presented at *Conference on Elements of Discourse Understanding*, University of Pennsylvania. (In Joshi, Weber, and Sag, 1981, 217–230.)

Perrault, C. R., and Cohen, P. R. 1979. Elements of a Plan-Based Theory of Speech Acts. *Cognitive Science* 3:197–212.

Perrault, C. R., and Allen, J. F. 1980. A Plan-based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics* 6:167–182.

Perry, J. 1977. Frege on Demonstratives. *Philosophical Review* LXXXVI: 474–497.

Perry, J. 1979. The Problem of the Essential Indexical. *Nous* xiii:3–21.

Perry, J. (in progress) Efficient and Benevolent Cognition. CSLI, Stanford University.

Peters, S., and Ritchie, R. (in progress) Phrase Linking Grammars. University of Washington, Department of Computer Science in conjunction with the University of Texas, Department of Linguistics, Technical Report.

Plotkin, G. 1972. Building-in Equational Theories. *Machine Intelligence* 7:73–90.

Pollack, M. 1984. Good Answers to Bad Questions. *Proceedings of the Canadian Society for Computational Studies of Intelligence.*

Pollard, C. 1984. Generalized Phrase Structure Grammar, Head Grammars, and Natural Language. Doctoral Dissertation, Stanford University, Stanford, Calif.

Putnam, H. 1975. The Meaning of Meaning. K. Gunderson (Ed.), *Language, Mind, and Knowledge*. Minneapolis: University of Minnesota Press.

Reichman, R. 1978. Conversational Coherency. *Cognitive Science*. 2:4.

Reichman, R. 1981. Plain Speaking: A Theory and Grammar of Spontaneous Discourse. Technical Report No. 4681, Bolt, Beranek, and Newman, Cambridge, Mass., Doctoral dissertation.

Reiter, R. 1978. On Reasoning by Default. *Proceedings of the Second Symposium on Theoretical Issues in Natural Language Processing.* Urbana, Ill.

Reiter, R. 1980. A Logic for Default Reasoning. *Artificial Intelligence* 13:81–113.

Roach, K. 1984. Mathematic Properties of Head Grammars. *Symposium on Mathematics of Language*. University of Michigan, Ann Arbor, Michigan. October

Rooth, M. E. 1985. *Association with Focus*. University of Massachusetts, Department of Linguistics, unpublished thesis.

Rosenschein, S. J., and Pereira, F. C. N. (in progress) Knowledge and Action in Situated Automata. SRI Artificial Intelligence Center Technical Note.

Rosenschein, S. J. (in progress) Natural Language Processing. In I. Benson (Ed.), *Intelligent Machinery: Theory and Practice*. Cambridge, England: Cambridge University Press.

Sadock, J. M. 1974. *Toward a Linguistic Theory of Speech Acts*. New York: Academic Press.

Schiffer, S. R. 1972. *Meaning*. Oxford: Oxford University Press.

Scott, D. 1973. Data Types as Lattices, *SIAM Journal of Computing* 5:522–587.

Scott, D. 1982. Domains for Denotational Semantics. In M. Nielsen, and E. Schmidt (Eds.), *Proceedings of ICACP 82*. Springer-Verlag, Notes in Computer Science 140:577–610.

Searle, J. R. 1969. *Speech Acts*, Cambridge, England: Cambridge University Press.

Sells, P. 1985a. Anaphora with Which. *Proceedings of the Fourth Annual West Coast Conference on Formal Linguistics*. Stanford, Calif.

Sells, P. 1985b. Restrictive and Non-Restrictive Modification. ms., Center for the Study of Language and Information.

Schank, R. C. 1975. *Conceptual Information Processing*. New York: American Elsevier.

Schank, R. C. 1982. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge, England: Cambridge University Press.

Schank, R. C., and Abelson, R. P. 1977. *Scripts, Plans, Goals and Understanding*. Hillsdale, N.J.: Lawrence Erlbaum.

Shankar, K. V., and Joshi, A. 1985. Computational Properties of Tree-Adjoining Grammars. Will be presented at *The American Association of Computational Linguistics Conference* to held in Chicago 1985.

Shepard, R. 1978. The Mental Image. *American Psychologist* 33:125–137.

Shepard, R. 1982. Perceptual and Analogical Bases of Cognition. In J. Mehler, E. Walker, and M. Garrett (Eds.), *Perspectives on Mental Representation*. Hillsdale, N.J.:Erlbaum. 49–68.

Shepard, R. N. and Metzel, J. 1971. Mental Rotation of Three Dimensional Objects. *Science* 171:701–703.

Shepard, R. N., and Cooper, L. A. 1982. *Mental Images and Their Transformations*. Cambridge, Mass.: MIT Press/Bradford Books.

Shieber, S. M., Uszkoreit, H., Pereira, F., Robinson, J., and Tyson, M. 1983. The Formalism and Implementation of PATR-II. In B. Grosz and M. Stickel (Eds.), *Research on Interactive Acquisition and Use of Knowledge*. SRI Final Report 1894, SRI International, Menlo Park, Calif.

Sidner, C. and Israel, D. 1981. Recognizing Intended Meaning and Speakers' Plans. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*. Vancouver, British Columbia, August. 203–208.

Sidner, C. 1981. Focusing for Interpretation of Pronouns. *American Journel for Computational Linguistics* 7:217–231.

Sidner, C. 1984. Focusing in the Comprehension of Definite Anaphora. In M. Brady and R. Berwick (Eds.), *Computational Models of Discourse*. MIT Press: 267–330.

Smith, B. C. 1984. Reflection and Semantics in LISP. Reports no. CSLI-84–8.

Smith, B., and des Rivieres, J. 1984. The Implementation of Procedurally Reflective Language. Report No. CSLI–84–9.

Stalnaker, R. 1980. A Note on Non-monotonic Modal Logic. Department of Philosophy, Cornell University, unpublished manuscript.

Stalnaker, R. 1984. *Inquiry*. Bradford Books.

Stampe, D. W. 1977. Toward a Causal Theory of Linguistic Representation. In P. A. French et. al. (Eds.), *Midwest Studies in Philosophy*. Minneapolis: University of Minnesota Press. 2:42–63.

Stucky, S. (in progress) The Use/Mention Distinction Reconsidered in Light of Some English Expressions. SRI International, Menlo Park.

Thomason, R. 1980. A Model Theory for Propositional Attitudes. *Linguistics and Philosophy* 4:47–70.

Thomason, R. 1983. Remarks on Mood and Conditionals. Department of Linguistics, University of Pittsburgh, unpublished manuscript.

Turner, R. 1983. Montague Semantics, Nominalization, and Scott's Domains. *Linguistics and Philosophy* 6:259–288.

Van Benthem, J. A. F. K. 1983. *The Logic of Time*. Dordrecht, Holland.

Warren, D. H. D., and Pereira, F. C. N. 1982. An Efficient, Easily Adjusted System for Interpreting Natural Language Queries. *American Journal of Computational Linguistics* 8:123–138.

Weber, B. 1978. A Formal Approach to Discourse Anaphora. Report 3761, Bolt, Beranek, and Newman, Inc., Cambridge, Mass.

Weber, B. 1981. Discourse Model Synthesis: Preliminaries to Reference. In A. Joshi, B. Weber, and I. Sag (Eds.), *Elements of Discourse Understanding*. Cambridge, Mass.: Cambridge University Press, 283-291.

Weber, B. 1983. Taking Initiative in Question Answer Systems. In N. J. Cercone (Ed.), *Computational Linguistics*. New York:Pergamon Press.

Wexler, K., Culicover, P. 1980. *Formal Principles of Language Acquisition.* Cambridge, Mass.: MIT Press. (There were also numerous articles but this book is the culmination of the research.)

Winston, P. H. 1979. Learning and Reasoning by Analogy. *CACP* 23:689–703.

Estimate of Needs

The healthy development of interdisciplinary basic research on Information, Computation and Cognition requires support for

- exchange of researchers
- postdoctoral fellowships
- student training grants
- electronic networking
- computing equipment
- research centers, group projects, and individual investigators

Basic research on Information, Computation and Cognition is being carried out in a spectrum of research environments organized according to widely varying models. Funding is needed for grants ranging from awards to an individual investigator, perhaps with an essential colleague, through group projects of various sizes, to research centers involving twenty or more investigators engaged in related projects and serving the wider research community in a variety of ways. Encouragement of collaborative work that crosses the boundaries of disciplines and of institutions, including the boundary between academe and the research arms of industrial institutions, is an essential requirement for the advancement of research in this area. Current cooperative arrangements between universities and industry often focus on applied research, but the need and potential for increased cooperation on basic research presents opportunities for federal money to be used with great leverage in stimulating new arrangements for collaborative inter-institutional research.

At present we estimate that this emerging interdisciplinary research area engages roughly five hundred (500) investigators at about fifty (50) sites in the United States (see "Description of Research Community," p. 72). Many of these investigators are unable, due to lack of resources or other constraints or commitments, to devote their full research time to basic investigations in the field of Information, Computation and Cognition; thus the full-time equivalent number of researchers in the area may be closer to two hundred (200).

In this section of the report, we attempt to estimate the area's support needs over the next five years. Our estimates are for direct costs only; no indirect costs are included in the figures. In our view, a first year sum of approximately $20 million in new funding is needed for a program to put basic research in Information, Computation and Cognition on a healthy footing. This total includes about $15 million in annual research costs plus $5 million for purchase of computing equipment. Over the subsequent five years the amount needed for operating costs would at least double, but the

requirement for acquisition of computing equipment could decrease to $1 million a year. The year one estimate breaks down as follows.

Capital Costs (in millions)	
Computing Equipment	$ 5.0
Operating Costs (in millions)	
Network Communications	$ 0.5
Faculty Development and Exchange of Researchers	1.875
Postdoctoral Fellows	0.75
Student Training	0.75
Other Research Costs	11.225
Total First Year Operating Costs	$ 15.1
Combined First Year Total	$ 20.1

Much of the needed basic research will likely be conducted at universities, and not-for-profit research institutions, under the financial sponsorship of the federal government. Industry can also help by providing both equipment and the time of researchers in industrial laboratories. The partnership between industry, government and academe is as appropriate for the area of Information, Computation and Cognition as it is in pharmaceuticals, where corporations participate in a variety of ways, including the direct gift of research grants, to support basic research in chemistry at universities. Applied research and development benefits from the results of basic research, and basic research benefits from the flow of problems, ideas and resources.

We now explain how we arrived at these estimates, beginning with items that are most naturally treated on a national basis.

The immediate capital needs for computing equipment we estimate at $5 million. If this amount is made available for acquisition of computing equipment in one year, then approximately $1 million may be sufficient each following year to allow for growth of the field. The $5 million initial figure arises from an estimate that the 50 sites where basic research in Information, Computation and Cognition is underway would need an average of $100,000 each for purchase of computing equipment. This average seems reasonable since sites where there is only one researcher or a very few might need only $15,000 to $50,000 for purchase of workstations. Locations having a group of five to ten researchers would need a time-sharing system, costing possibly $250,000; however, there are fewer of these sites. And the small number of sites with very large research groups will need local area computer networks, but would be likely to have a significant amount of computing equipment available already.

When the field of computer science began, the magnitude of the equipment needs its research entailed was not adequately foreseen. Consequently, a situation persisted for a number of years in which computer science departments with sizable funding from the Department of Defense were adequately equipped for research, while others experienced very great difficulties in obtaining essential equipment and were essentially unable to pursue experimental work. We think it important that the field of Information, Computation and Cognition not suffer this fate, with attendant segregation into privileged and underprivileged researchers. Computational facilities are as essential a resource as library materials for basic research in Information, Computation and Cognition; and since research institutions do not provide computing as part of overhead, it is vital that research grants pay for it as a direct cost.

One-Time Capitalization of Computer Equipment

50 sites @ $100,000 = $5.0 million
In subsequent years it could decrease to
approximately $1 million per year.

Ideally, gifts of equipment from industry could meet a significant part of the needs of this research. This would require a change in current giving patterns, however, as major grants of computing equipment for research have tended to go primarily to computer science departments. Basic research in Information, Computation and Cognition is inherently interdisciplinary, and departments of linguistics, philosophy, and psychology are likely to be as heavily involved as departments of computer science. Until prospective donors change patterns and begin giving significant amounts of equipment to interdisciplinary research projects, the federal agencies committed to support basic research in the field will have to be prepared to play a substantial role in fulfilling the concomitant equipment needs.

Electronic network communications between researchers in Information, Computation and Cognition, on a nationwide basis as well as within their own local institution, is essential to facilitate the necessary cross-disciplinary interactions. Researchers need access to computer networks such as the Computer Science Network, or the Cognitive Science Network if that is developed. This entails their being able to pay the costs of electronic network communication, which probably average $1,000 a year per researcher.

Network Communications

500 @ $1,000 = $0.5 million

There is a real need for faculty development and exchange of researchers in Information, Computation and Cognition, to provide senior researchers (including scientists at academic and nonacademic research institutions)

opportunities to gain further training and research experience across disciplinary and institutional lines, to allow for closer contacts with colleagues at other institutions for the purpose of collaborative research, and to make greater time available for research. New researchers now versed in only one of the core disciplines may also be drawn into the interdisciplinary area by opportunities to learn one or more others. Half-salary support for fifty such appointments a year seems appropriate. (For present purposes, we assume an average academic-year salary of $45,000, annualized for convenience of calculation.)

Faculty Development and Exchange of Researchers

50 @ 1/2 × $60,000 plus 25% benefits= $1.875 million

Postdoctoral fellowships are an excellent means of channeling bright young Ph.D.s who are already well trained in one or two of the core disciplines into this interdisciplinary area through training and research experience in another aspect of it. Experience indicates that on a nationwide basis, at least forty highly-qualified postdoctoral scientists would benefit from such training. Assuming two-year fellowships for such scientists, we have phased in twenty in the first year of an initiative, expanding to forty in the second year. We envision that some funds for postdoctoral fellowships could be dedicated to a national competition, the chosen fellows to hold their fellowship at a site of their choosing, while other fellowship funds could be awarded to institutions, allowing them to select postdoctoral scholars.

Postdoctoral Fellows

40 @ $30,000 × 1/2 in first year + 25% benefits = $0.75 million
40 @ $30,000 in second year + 25% benefits = $1.5 million

Training grants for graduate students are an important priority, especially in view of the longer than average training (as much as five years or more) required due to the interdisciplinary nature of the field. Providing four-year support for two hundred students, one for each full-time-equivalent current faculty member, should assure the area of a reasonable supply of trained scientists in the future. These could be phased in gradually, one-fourth the first year and increasing each following year.

Student Training

200 @ $15,000 × 1/4 in first year = $0.75 million
200 @ $15,000 by fourth year = $3.0 million

Other support needs of basic research in Information, Computation and Cognition are less easily estimated on this basis. We believe, however, that factoring out the nationally estimatable needs just discussed facilitates assessment of the balance, which comprise salaries of researchers, technical

and support staff, computer operations and maintenance, other experimental equipment, travel, workshops, support of researchers visiting centers and groups, materials and services. Estimates of the amounts needed for each of these categories must take into account the variation in distribution of needs from one institutional setting to another. For instance, one institution's scientists may find it best to organize their research activities into project groups that apply for grants separately, while another's scientists may obtain better effects from a single grant to a research center that encompasses a number of coordinated projects involving the researchers in various overlapping combinations. At other institutions, there will be just a single group project, while still others have only individual investigators engaged in basic research on Information, Computation and Cognition. Some centers may have large numbers of short- and longer-term visitors, serving as national and international points of information exchange among researchers. Scientists who are alone in the field at their institution may tend to have few or no visitors—though we will suggest a way to counteract this tendency.

To put a dollar figure on the size of need, one must hypothesize some sort of budgetary analysis. In view of the variation in circumstances from one institutional setting to another, this is inevitably a somewhat speculative exercise. The assumptions and figures that follow are not in any way intended to direct the shape of proposals for, or awards of, support. They should be seen as just what they are: hypothetical cases that may be useful in estimating the support required by a sizable and varied community of researchers whose particular needs in each instance may vary considerably from the hypotheses considered here.

For purposes of making estimates, we consider two cases: that of an individual researcher applying for a grant to support basic research in Information, Computation and Cognition who is perhaps the only individual at his or her institution engaged in this field, and that of a group of researchers applying for support of their research whether as a single project or as a range of projects coordinated within the framework of a research center. Our estimates suggest that the cost per investigator does not differ as much as might be expected across these cases.

With these preliminaries out of the way, let us first consider what could be common to the two cases. Investigators could generally be expected to request

- partial salary for research time,
- student research assistance
- employee benefits
- networking costs

- computer operating costs
- travel
- materials and services

Some might also request support for a postdoctoral scientist to work with them.

A lone individual may well differ from a member of a group or research center, in that the latter would be more likely to need

- technical and support staff
- workshops and visitors

On the other hand, we suggest that a lone investigator might particularly benefit from the interaction that could be provided by having another researcher visit him or her on a faculty development or research exchange leave. Therefore preference should be given to proposals that would enable researchers to spend their visiting time at institutions where otherwise a basic researcher in Information, Computation and Cognition would be isolated.

Of the items just enumerated, we have treated support for students, networking, postdoctoral fellows, and faculty development (researcher exchange) separately above. Thus at this point, only the other items remain to be estimated.

The amount of salary support for research time that is appropriate will vary considerably from grant to grant. A lower bound is provided by two months summer support. While the upper bound would be twelve months support (for instance, for a researcher at a not-for-profit research corporation), this would not be a typical case. A more frequent case might be support for half the academic year and two summer months. Because of the great time demands in interdisciplinary research imposed by the need to communicate across disciplinary boundaries, it is desirable to support researchers for more, rather than fewer months per year. Thus in estimating need we assume half support from research grant funds for the investigator. This assumption is in force for lone investigators and for members of groups or research centers. We have also annualized salaries, although academic salaries are customarily stated on a nine-month basis. Given the diversity of salaries across rank, department, and institution, it is impossible to predict precisely what average figure will obtain in academic 1986–87. In our calculations, we use an academic-year salary of $45,000, annualized to $60,000, as being a reasonable estimate. This corresponds to $10,000 for two months' summer salary.

After one other explanation, we can perform some illustrative calculations. It seems likely that the main source of variation in cost per investigator of grants to project groups and research centers will be the number of visitors from other institutions hosted by the group or center. Larger

centers will naturally tend to perform more of this function in service to the field as a whole. As the number of visitors increases, they necessitate staff to support them. The non-linearity in cost per investigator as group size varies is caused mainly by this fact, and is in fact quite small. If a group of ten researchers budgets $100,000 annually for visitors and workshops, and a research center with twenty investigators requires $250,000 for this purpose due to the added staff needed to handle the greater number of visitors, then the cost per investigator is only $2,500 higher at the research center. Variation like this is clearly insignificant in estimates such as we are making. Thus we use the ten investigator group in our illustrative calculation.

Consider, finally, the mix of research centers having twenty or more investigators, project groups having five to ten investigators and individual investigators. We are recommending that two or more major research centers should receive funding to assure the health of the field; thus it is conservative to assume that forty (40) scientists will be supported at such institutions. We assume that eight project groups with an average size of seven and one-half investigators will be supported at other institutions, or sixty (60) more scientists receiving support. And we assume that another fifty (50) scientists will receive support as individual investigators. Note that this would be sufficient to provide some support to each of the approximately fifty institutions where basic research in Information, Computation and Cognition is currently underway.

Now to the calculations of other costs per investigator. For a group project involving ten researchers, our sample calculation is as follows.

Research salaries	10 @ 1/2 support × $60,000 (12 month salary)	$ 300,000
Technical and support staff	3 @ $40,000	120,000
Employee benefits	$420,000 @ 25%	105,000
Visitors and workshops		100,000
Computer operations		50,000
Experimental equipment		50,000
Travel		50,000
Materials and services		50,000
TOTAL excluding networking, postdocs, students and development		$ 825,000
	Per Investigator	$ 82,500

We saw that the cost of these items per investigator in a twenty investigator research center might be $2,500 higher, or $85,000.

Now we estimate the cost of a lone investigator. In these fifty cases, we include the cost of faculty development/exchange researchers for purposes of calculation since we are suggesting that preference be given to these leaves being taken at institutions where a basic researcher on Information, Computation and Cognition would otherwise be isolated.

Research salary	1/2 support @ $60,000	$ 30,000
Exchange researcher	1/2 support @ $60,000 (12 month salaries)	30,000
Employee benefits	$60,000 @ 25%	15,000
Computer operations		5,000
Experimental equipment		5,000
Travel		5,000
Materials and services		5,000
TOTAL excluding networking, postdocs and student(s)		$ 95,000

These figures now let us estimate the requirements for other research support not dealt with earlier on a national basis. Fifty individual grants averaging $95,000 for the items considered here account for $4.75 million. Sixty investigators supported on group projects at an average of $82,500 for the items estimated come to $4.95 million. And forty investigators supported at centers with an average cost of $85,000 for these items amount to $3.4 million. These figures total $13.1 million, but include $1.875 million for all fifty exchange researchers. So the total for needs not estimated earlier on a national basis is $11.225 million.

In summary, in the first year of a program the following new support is needed for basic research in Information, Computation and Cognition.

Capital Costs (in millions)		
Computing Equipment		$ 5.0
Operating Costs (in millions)		
Network Communications		$ 0.5
Faculty Development and Exchange of Researchers		1.875
Postdoctoral Fellows	(in first year)	0.75
Student Training	(in first year)	0.75
Other Research Costs		11.225
Total First Year Operating Costs		$ 15.1
Combined First Year Total		$ 20.1

The above estimates of support needed for basic research in Information, Computation and Cognition concentrate on types of activity to be supported; our calculations may not correspond closely to a budget that any particular researcher or collection of researchers would be likely to submit. It may be of interest, therefore, to look at the estimated $15.1 million first-year operating costs from another viewpoint. We may assume that major research centers might receive support levels of $2.5 million to $5 million, group research projects support of $0.5 million to $1.0 million, and individual investigators levels of $50,000 to $125,000. If this is correct, then as many as eight group research projects and up to fifty individual investigators could be supported for approximately $7.9 million. Supporting the two or more major research centers we are recommending might cost $7.2 million.

Clearly the first-year total of $15.1 million will need to increase in succeeding years. The level of basic research activity in the field can be expected to grow with the number of investigators, postdocs and students. Over a five-year period, we estimate that the total research need will probably at least double. It is hoped that industrial as well as federal sources of support will become available during that period to help meet the need.

Note that all estimates are for direct costs only. None include any indirect costs, which we have not attempted to estimate as they vary considerably from one institution to another. Nowhere are they negligible, however. The $20.1 million we are recommending as a beginning program level is based on a conservative, even low, estimate of needs for support of basic research in Information, Computation and Cognition. However, given the scientific and technological significance of this work, an investment of at least $5 million to $10 million more than we are recommending may be called for.

Description of Research Community

Interdisciplinary research in Information, Computation, and Cognition is pursued at a number of institutions around the United States. These include both universities and industrial laboratories. The size and disciplinary mix varies considerably from place to place, as do the specific problems under investigation. Thus, even though the diverse research groups share a number of assumptions and goals, there is no straightforward way of identifying which institutions have them or which individuals participate in them. The following data are consequently largely impressionistic, having been pieced together from the best guesses of a handful of researchers.

We estimate that the total number of investigators engaged in research of the sort described here is on the order of 500 to 1000. This number excludes students and is limited to people working in the United States.

The following is a list of institutions where interdisciplinary groups of scientists are actively pursuing research on Information, Computation and Cognition. It is not comprehensive, but it does include all of the largest centers of activity.

AT&T Bell Laboratories
Bolt Beranek and Newman Inc.
Boston University
Brandeis University
Brown University
Carnegie-Mellon University
Columbia University
Cornell University
City University of New York
Dartmouth College
Fairchild Research Laboratory
Florida State University
Hampshire College
Harvard University
Hewlett-Packard Laboratories
IBM Thomas J. Watson Research Laboratory
Indiana University
Massachusetts Institute of Technology
Ohio State University
Pennsylvania University
Princeton University

Rutgers University
SRI International
San Jose State University
Stanford University
Tufts University
University of Arizona
University California, Berkeley
University California, Irvine
University California, Los Angeles
University California, San Diego
University California, Santa Barbara
University California, Santa Cruz
University of Chicago
University of Colorado
University of Connecticut
University of Illinois
University of Massachusetts, Amherst
University of Michigan
University of Minnesota
University of North Carolina
University of Pittsburgh
University of Rochester
University of Southern California
University of Texas, Austin
University of Washington
University of Wisconsin
Vanderbilt University
Xerox PARC
Yale University

Estimate of Currently Available Resources

Because of the interdisciplinary and relatively novel character of the research under discussion, it is difficult to obtain reliable data regarding either the levels of funding currently available or the amount of potentially important research that goes unfunded. This section contains the information we were able to assemble on these subjects. While our numbers are by no means authoritative, they make one point very clearly: basic, theoretical research on Information, Computation and Cognition is extremely hard to fund. That is, the applications for support of such research vastly exceed the funding currently being provided. It is also our collective impression, and certainly our experience, that submitted applications substantially underrepresent the willingness of researchers to undertake interdisciplinary basic research in this area. There is, we believe, a widely shared perception that apart from the initiatives encouraged by the Sloan Foundation and the System Development Foundation in recent years, large-scale long-term interdisciplinary projects in basic research in this area are extremely difficult to fund. This leads many investigators to concentrate their proposal-writing efforts on smaller-scale narrowly defined projects and/or on projects with a heavy proportion of applied work of immediate use to industry or defense. In some of the contributing disciplines such as logic and philosophy, external funding appears to be extremely difficult to obtain at all.

I. Sample Institutions

We have compiled data from four major universities where high quality research of the sort we wish to encourage is going on (Carnegie-Mellon University, the University of Michigan, the University of Pittsburgh, and the University of Texas), and two of the major private research organizations (Bolt, Beranek, & Newman and SRI International). These are just six of fifty institutions listed in the preceding section where basic research on Information, Computation and Cognition is underway. The total requests by these six institutions come to $14,756,000 per year. The total awards come to $3,896,000 per year, barely one quarter of the amount requested. This probably overestimates the overall success rate, because 1) non-educational institutions do not ordinarily apply for large grants which they feel they cannot compete for successfully, and 2) the figures above include projects that have a significant component of theoretical research on relevant topics (recognizing that the question of what counts as theoretical research is somewhat subjective), but no attempt has been made to subtract any-

thing for what does not come under the proposed initiative. Nevertheless, the figures show that, even at the top research institutions in the country, projects in the areas described are unlikely to be funded. The figures also include approximately $1,000,000 one-time only gifts from the System Development Foundation. As the seed money from the System Development Foundation and the Sloan Foundation is exhausted, prospects for proposals in Information, Computation and Cognition will become even bleaker.

II. Potential Sources of Government Funding

A. Department of Defense

Four agencies within the Department of Defense were contacted and asked how much basic research in Information, Computation and Cognition they support, and what level of funding requests they receive. The following summarizes the responses.

Air Force Office of Scientific Research: The gross figure for AFOSR's AI budget in FY 85 is $2,281,171. This amount has been reduced by 20 percent for FY 86, as part of a decision to move the computer science program in the direction of computer science aspects of large scale scientific computation and to reduce funding in AI. Even prior to this reduction, no more than one quarter of AFOSR's AI funding was for interdisciplinary theoretical research on Information, Computation and Cognition.

Office of Naval Research: Requests for basic research in artificial intelligence for FY 85 totalled $2,905,000. Of that, $869,000 was actually funded. Most of this research was on expert systems and natural language processing, and it did not include work in robotics. An additional $400,000 went to interdisciplinary work in computer science and psychology. No more precise information was available on how much of the research sponsored by ONR really fit into the subject areas of this report.

Defense Advanced Research Projects Agency: DARPA's FY 86 budget for theoretical computer science is $8 million. We were not able to obtain a breakdown regarding how much of this goes for theoretical and interdisciplinary research, nor do we know how much more support was requested.

Army Research Institute: ARI has a total budget of $5.5 million, but the sort of research we described was not really central to its concerns. Most of the artificial intelligence research it funds is mission-oriented.

In summary, hard data are not readily available regarding levels of available DoD funding for research on Information, Computation and Cognition. What reliable information we have indicates that basic interdisciplinary research in these areas receives very little support from DoD. Personal expe-

rience and anecdotal evidence indicate that this is a fairly general pattern.

B. National Science Foundation

The following table summarizes funding levels in relevant NSF programs over the past three years. These figures are estimates of how much each program spent on research falling within the interdisciplinary area of Information, Computation and Cognition that we have identified.

	'83	'84	'85 (Est.)
IST	3.5M	4.0M	4.5M
COMP RES	1.2M	1.4M	1.7M
LING	639K (13 awards)	618K (12 awards)	900K[1]
MEM & COG	1.128M (15 awards)	1.675M (23 awards)	1.2M[2]
TOTALS	6.467M	7.693M	8.3M

[1]To date 458K, 8 awards.
[2]To date 213K, 4 awards.

III. Conclusions

The discrepancy between these totals above and the calculations of what is needed to sustain the Information, Computation, and Cognition community is evident. The rate of increase in NSF funding is not nearly sufficient to fill the gap created by the anticipated cessation of the Sloan Foundation and System Development Foundation funding (which has increased from approximately $1 million in 1977 to over $10 million in 1985). Recommendations by Eric Wanner (Sloan) and Carl York (SDF) for a federal initiative make this clear. (See the section, "History of Alfred P. Sloan Foundation and System Development Foundation Support.") Existing NSF grants, moreover, are for individual researchers or small groups, averaging $50,000 to $75,000 per year. They are not adequate for the sort of larger scale collaborations that have been fostered by the private foundations. Given private industry's understandable emphasis on applications and the Defense Department's low level of support for interdisciplinary theoretical research, dramatically increased NSF support is crucial for the continued robustness of basic scientific research in Information, Computation, and Cognition.

It seems evident from these figures that even at high-quality institutions, funding in these areas is low in proportion to need. Several less obvious observations about the current funding situation need to be mentioned:

A. Theoretical research funding is very low in proportion to applied, product-oriented funding. Under the latter sort of funding, even researchers with strong theoretical interests will find it difficult to expend much time on basic research.
B. Long-term theoretical funding is scarce. This makes it difficult to create stable theoretical research groups.
C. Under this funding pattern, it is difficult to obtain funding for post-doctoral researchers to pursue theoretical projects. It is especially hard to fund such researchers if their graduate training was not computational.

Alfred P. Sloan Foundation and System Development Foundation Support

The Sloan Program in Cognitive Science

Eric Wanner, Alfred P. Sloan Foundation

The funding program in cognitive science undertaken by the Alfred P. Sloan Foundation will, by the time of its conclusion in 1987, have expended approximately $17.4 million (for direct costs) over a period of some 10 years. In its initial phase in 1977–78, the program principally supported integrative activities such as workshops, conferences, and visiting scientists. This phase of the program consisted of a large number of relatively small grants designed to foster increased communication and collaboration across the boundaries of the component cognitive sciences, which remained quite separate (and even hostile) in the mid-seventies. By 1979, the program entered a second phase in which training grants of $435,000 each were awarded to 12 universities in order to underwrite (partially) both pre- and post-doctoral training programs in cognitive science. In 1981, the Sloan program initiated its final phase with a group of 9 grants, ranging between $435,000 and $2.17 million, for the development of research and training centers in cognitive science. These centers are located at MIT, UC-Berkeley, Stanford, Pennsylvania, Carnegie-Mellon, UC-Irvine, Texas-Austin, Cornell, and Rochester. All of them continue to operate successfully. There is a good deal of diversity across these centers both in the substantive focus of the research and in the types of activities supported. However, in general, Sloan funds have been expended for faculty development, computational facilities, seed money for research, pre- and post-doctoral fellowships, visiting scientists, and even undergraduate curriculum development.

The Sloan program was conceived and implemented as a pilot project designed to help the field of cognitive science coalesce conceptually and to bring together sufficient concentrations of talent from the component fields to permit progress on the very difficult problems posed by the analysis of intelligence. These goals have now been largely realized, but the ultimate success of the Sloan initiative will depend upon subsequent developments. Clearly, the field must find a reliable source of support over a sufficiently long period to permit sustained progress. In the view of the Sloan Foundation, such support would ideally be targeted for theoretical work on basic questions, since much of the funding now available (both from government agencies, such as DARPA and NIE, and from the burgeoning 'cognitive' in-

dustry) is applied in nature. Furthermore, it would also be optimal in our view if support were concentrated in a relatively small number of centers where interdisciplinary cooperation is possible and where sufficient concentrations of talent now exist. To estimate the scale of such centers, it might be useful to consider a continuum of size with Sloan's nine centers at an average annual budget of about $300,000 anchoring the low end of the scale and SDF's Center at Stanford pushing the upper end with an annual budget of $4–5 million. Although the Sloan Centers functioned adequately on minimal budgets, it is probably true that we tested rather severely the limits of austerity in this regard. In retrospect, we believe that funding levels at least twice the Sloan level would have been more appropriate. Moreover, as the field matures and work becomes focused on a small set of critical problems, it may well be the case that larger centers would be warranted. Therefore, something roughly midway between the Sloan level and the SDF level of budget might produce the ideal operating environment for future centers.

A Brief History of the System Development Foundation's Program in Computational Linguistics and Speech

Carl M. York, System Development Foundation

The System Development Foundation received its first proposal in the area of computational linguistics and speech in early 1982. In the summer of 1984 the Foundation had committed most of its funds and decided not to accept any new proposals. During that period the Foundation made grants totaling $78 million, and of that amount, $26 million went into the area of computational linguistics and speech. This substantial financial commitment reflects the importance placed on this area of research by the Trustees and Staff of the Foundation.

The Foundation's view of this field of research is reflected in the following excerpt from the Foundation's 1983 Annual Report:

> *Computational linguistics and speech* is a field exploding with ideas. For example, viewing the generation, interpretation, and acquisition of language as a computational process provides new constraints for language research. It takes the *deus ex machina* out of linguistic theory and forces the investigator to explain in empirically verifiable detail *how* meaning is extracted from surface structure. Similarly, viewing computation as a linguistic process appears to offer important insights into the nature of computation that have direct implications for practical programmers. It now appears that a common semantics may underlie both natural and computer language.

> Despite the flood of ideas, support in this area has been directed primarily toward the engineering of applied systems, rather than an understanding of fundamental principles.

In reviewing proposals that came in from diverse institutions and researchers in this broad area of investigation, several features emerged. First, the individuals who were interested in the topics in this field came from widely disparate disciplines and backgrounds. Philosophers, logicians, mathematicians, linguists, computer scientists, phoneticians, cognitive scientists, and clinical speech pathologists came forward with proposals for research related to the general area. The sizes of these proposals fell rather naturally into three categories, depending on the number of collaborators involved and the kind of operating organization that was proposed. These three categories were:

- large centers,
- multi-investigator projects, and
- individual principal investigators.

It would have been possible to separate the large centers into multi-investigator projects; however, the SDF Board of Trustees decided that a key criterion for funding should be a demonstrated interaction between university and industrial researchers. They reasoned that a center located on a university campus would provide for maximum interaction between industrial researchers and academics and that this in turn would increase the rate at which research ideas could be transformed into practical products. It was felt that joint university-industry projects would tend to focus and isolate the individual researchers by tying them to specific goals, while a center would provide a more flexible management structure and a more stimulating research milieu.

In reviewing possible locations for a large center, only four clear candidates were found which had credible interdisciplinary activities and links to industry. These institutions were Stanford, in the heart of "Silicon Valley," M.I.T., with its "Route 128" connections, Carnegie-Mellon/Pittsburgh, and the University of Texas at Austin, with the rapidly growing MCC. The primary reason for choosing Stanford over the other possible candidates was that a diverse group had already come together, was meeting regularly, and was beginning to produce ideas and publications. The other candidates were judged not to be as far along in their development.

SDF's limited funds precluded funding more than one such center. Detailed descriptions of the Stanford Center for the Study of Language and Information (CSLI) are available. In its third year of operation, CSLI has a staff of about 70 and an annual operating budget of about $5 million a year.

At the multi-investigator project level, SDF received approximately 20 proposals with requests for support of three or four faculty and five or more graduate students. The requested level of funding ran about $100,000 per faculty person. Industrial collaboration was not emphasized in the six such projects that were eventually funded by the Foundation.

Finally, the third category of individual researchers led to just over 40 proposals with funding requests in the range from $50,000 to $150,000, depending on the project and equipment needs. SDF funded only four proposals in this category, because a determined effort was made to concentrate the Foundation's funds on the larger interdisciplinary efforts.

Table I summarizes the data on SDF proposals which were received and the number of those that received funding. The data are separated by the size of the proposal into the three categories discussed above.

TABLE I

SDF PROPOSALS BY THEIR SIZE

	Proposals Received	Proposals Funded
"Centers" (>25 staff ≥ $2M/yr)	4	1
Projects (>3 staff ≈ $250K/yr)	20	6
Individuals	40	4

Table II provides a financial overview of the proposals received by SDF. The funding that was requested is broken down into "Unfunded" and "Funded" portions, and then added to give the "Total." To give some idea of the annual dollar volume of requested funding, the next row in the table gives the average of dollar support requested, divided by the number of years of support requested.

TABLE II

SDF PROGRAM IN COMPUTATIONAL LINGUISTICS AND SPEECH

	Unfunded	Funded	Total
Funding Requested	$ 44.3M	$ 26.1M	$ 70.4M
Average $s Requested/Yr.	$ 12.4M	$ 8.3M	$ 20.7M
No. Faculty	114	59	173
$s/Fac./Yr.	$ 109K	$ 141K*	—

(*includes equipment)

Note: SDF grants do *not* include any institutional overhead

The number of faculty members who wished to be involved in the program was determined by counting those named in the various proposals.

Since these 173 individuals actually participated in writing proposals to SDF, this must be considered to be a lower limit on the actual number of persons interested in this interdisciplinary area of research. For internal planning purposes, SDF staff assumed that there were about three times this number of relevant researchers in the academic world and that the number would grow rapidly as the area became better understood.

The final row gives the dollars per year per faculty for funded and unfunded projects. The proposals tended to underestimate the amount of computing equipment required to do the work, and when SDF staff discussed the proposals for funding with the applicants, the cost went up significantly. It should also be noted that since SDF grants do not include institutional overhead costs, the funded figures should be increased by about 35% to 50% to get an estimate of the actual project cost. It should be noted that these costs per faculty member are comparable to the aerospace costs for "MTS" personnel.

There are at least two indicators that the SDF program has established the United States' leadership in this new field of basic research. Both indicators are based on an assessment of, and a comparison with, the Japanese efforts in this field. John Seely Brown and Brian C. Smith of Xerox report that, based on their visit to Japan in the Fall of 1984, the Japanese are certainly going to exploit the commercial potential of intelligent systems, but that they have not come to grips with the fundamental limits of their present mode of operation. These limits, and the need for a basic research program to explore them, were clearly perceived in the United States ten years ago. CSLI at Stanford was a direct outgrowth of the combined efforts of academe and industry to attack these fundamental problems. In a recent issue of *Time* (March 11, 1985), Dr. Tom Anyos of the Technology Group in Atherton, California, comments that:

> Although in the computer fields, especially on the software side, the Japanese have seemed to lag behind the Western World, recent advances in both hardware and software (often very specific to Japanese usage) portend a vigorous growth in this sector as well.

Both of these sources reflect the fact that the United States has a leadership role in this area of basic research underlying computer software.

The Foundation's program in computational linguistics and speech has focused its resources to enable a new field dealing with Language, Computation and Information to become established and set new directions for basic research which will be needed for a Sixth Generation of Computing. Because SDF will not be able to provide funds for this work after 1987, this leadership position is in jeopardy.

Role of Industrial Research Centers

Industrial research centers, by their very nature, offer an environment where investigators in the core disciplines are in close enough contact to interact productively. There are certain advantages to this setting and also inherent difficulties. The advantages include the availability of resources, computing facilities in particular, and the stability to pursue long-range goals without expending substantial time and effort on securing and maintaining necessary funding. While direct opportunities to interact with students are somewhat limited, adjunct professorships are readily available, and both co-op and postdoc programs are common. However, such programs do not allow sufficiently sustained contact to overcome inadequacies of traditional narrowly-focused graduate programs.

A more serious impediment to conducting basic research in the core disciplines in an industrial setting, however, is the competing demand of involvement in technology transfer. Both of these activities are usually represented in industrial research groups engaged in the area in question, but the mix varies widely from company to company, from person to person, and from time to time. This tremendous diversity of experience at various industrial research centers makes it difficult to generalize, but some of the more common problems include the following:

(1) Individuals who are free to spend all or even most of their time on basic research in the core disciplines are rare. An 80%–20% mix of applied research (often product-oriented) to basic research is typical.

(2) Choice of an underlying model is often influenced more by short-term performance considerations than is healthy for achieving significant long-term results.

(3) Over time, the general expectation is that basic research will lead to the development of a salable product. This leads to increasing pressure on time available for basic research.

(4) It is often the case that some of the best individuals and groups are the ones with the most responsibility for product development, leaving less qualified ones freer to conduct basic research.

(5) Estimation of technology transfer/basic research ratios at levels such as 80% probably overestimates the time available for research for several reasons; extraordinary demands on time, which are in fact quite ordinary, squeeze out research more than technology transfer. Furthermore, the specific demands of product development color the choices made in conducting basic research.

It should be noted that a very small number of industrial research settings do allow researchers the freedom to concentrate on basic research without the constraint of short-term, product-oriented goals; these settings are free of most of the impediments listed above. Such laboratories have traditionally provided a haven for interdisciplinary research. However, the lack of students with appropriate interdisciplinary backgrounds is a major frustration to researchers within these laboratories. While co-op programs are available, the limited contact with students is not sufficient for such programs to substitute effectively for appropriate university training.

For all these reasons, it is apparent that contributions to basic research in the core disciplines by groups from industry cannot be expected to constitute a dominant fraction of the work which will be carried out in the foreseeable future. Nevertheless, it is greatly to be recommended that industrial researchers be directly included in some fashion in the research institutions that may be set up to carry out that work. This provides leverage of research from industry, and also provides a source of concrete problems which often suggest new directions for basic research. It also encourages industry to contribute to the financial support of those research institutions.

Expository Papers

Language, Mind, and Information

John Perry

There are three points I wish to make.

First, the pick-up of information and the flow and transformation of information (and misinformation) across physical, psychological, linguistic, computational, and other sorts of information-carrying events and processes form a unified subject matter which intersects a number of disciplines, including linguistics, psychology, computer science, artificial intelligence, philosophy, and logic. We call this field *Information, Computation, and Cognition.* I shall sometimes shorten this to *information and intelligence.*

Second, mathematical theories of informational content, which are a crucial unifying tool for research in information and intelligence, are neither trivial nor impossible; such theories are emerging from a renaissance in logic that is already well underway.

Third, research into information and intelligence is underway in a number of interdisciplinary settings around the nation, and these activities are a valuable national resource. They have been aided by NSF and other agencies and in major ways by two private foundations, the Sloan and System Development Foundations; but funding that is more generally available and available on a continuing basis is needed.

The Flow of Information

Consider the cartoon strip on the next page. I want to use this to introduce two basic ideas: *(informational) content* and *the flow of (informational) content.*

In the first cartoon the dog Jackie has a broken leg. This situation or fact makes the statement at the bottom true. Now, the way I am using content, a number of different statements, using a number of different sentences, could have the same content as this statement. For example, one might point at Jackie and say "That dog has a broken leg." Jonny could say what he later thinks, "My dog has a broken leg." The vet could, as he does, say "She has a broken leg." All these statements say of the same object that it has the same property; they say of the dog Jackie that she has a broken leg. They do so in different ways, using different sentences; but they are all true in exactly the same case, namely, if Jackie's leg is indeed broken.

The simplest examples of the flow of information involve this constancy of content amid compensating changes in context and linguistic expression. I say, "I am from Nebraska." You say to me, "Oh, you are from Nebraska."

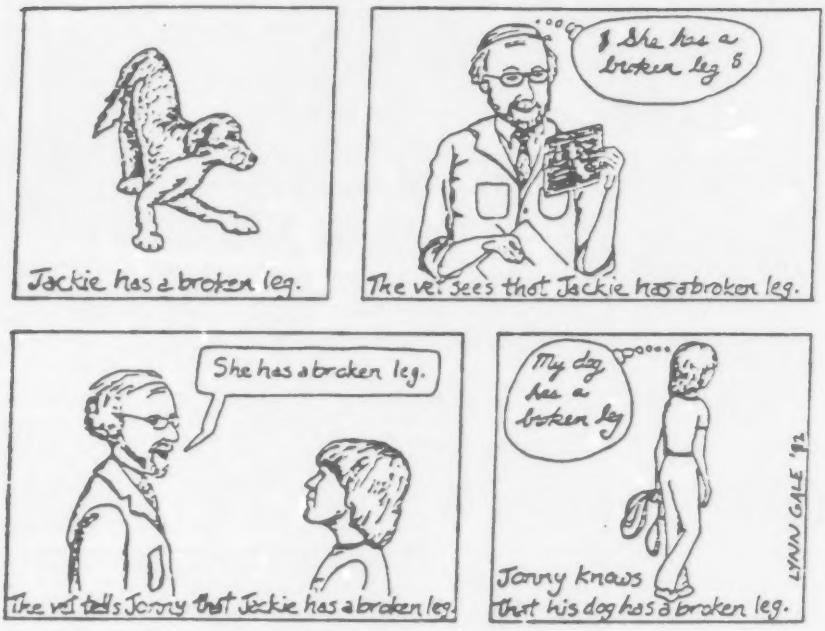

Jackie has a broken leg.
§ She has a broken leg §
The vet sees that Jackie has a broken leg.
She has a broken leg.
The vet tells Jonny that Jackie has a broken leg.
My dog has a broken leg
Jonny knows that his dog has a broken leg.
LYNN GALE '92

Figure 1. The flow of information

Later you point to me and say to someone else, "He is from Nebraska." Different linguistic events, in different contexts with different sentences employed, communicated the same content.

In our cartoon strip, we have a more complex and more interesting example of the flow of informational content. (Note that we cannot simply say "information" because strictly speaking information ought to be true, but misinformation can flow quite as easily as information—perhaps more easily.)

Here not just modes of linguistic expression, but the whole nature of the situations changes. We could note some more members of the sequence.

1. The x-ray *shows* that Jackie has a broken leg.
2. The vet *sees* that Jackie has a broken leg.
3. The vet *knows* that Jackie has a broken leg.
4. The vet *says* that Jackie has a broken leg.
5. Jonny *hears* that Jackie has a broken leg.
6. Jonny *knows* that Jackie has a broken leg.

7. After he enters the facts, Jonny's Macintosh *stores the information that* Jackie has a broken leg.

In the first, we have an x-ray—an inanimate structured object—which has a certain content: it *shows* that Jackie has a broken leg. Then we have the vet seeing that Jackie has a broken leg, a very complex process involving the x-ray, the eyes of the vet, and internal events. Knowing is also a state of the vet, due to states and processes of his central nervous system; but there is a difference between *seeing that* and *knowing that*. The vet will know that Jackie has a broken leg long after he has ceased to see that this is so. Then we have a linguistic event involving the production of a structured object, a sentence; since we are taking this to be a case of transmission of information, we assume that this production is caused by the knowledge: the vet said that Jackie had a broken leg in part because he knew that this was so. And so on.

I shall call all of these properties "content-properties." I shall call the recognition of such properties semantic classification, and the purposeful production of activities and artifacts to be semantically classified, semantic activity. (The word "semantics" is from the Greek "semantikos," for meaning.) When I read the newspaper and see that it *says that* Iran has invaded Iraq, this is a case of semantic classification of the words I see in the paper. When I spot Jerry Fodor's distinguished but deeply furrowed brow and I note that this philosopher *fears that* I am going to draw too close of an analogy between x-rays and minds, this is also a case of semantic classification. My giving this talk is a case of semantic activity. To understand me, you will have to recognize the sounds and words I use and the structure of the sentences that I produce. (Perhaps you will also recognize various stylistic deficiencies and mispronunciations.) But my goal is not your recognition of any of these properties of the noises I am making, but rather your recognition of their content. Later on, when I try to give a mild taste of how logicians have conceived of semantics, I fear the great logician Dana Scott sitting here to my left will bury his head in his hands. I will surely recognize that he *believes that* my oversimplifications threaten to distort a grand subject. That will be semantic classification of his state of mind on my part. If Professor Scott does this voluntarily, with the intent of communicating his view, that will be semantic activity on his part. If it is just an involuntary reaction to a distressing situation, of course, then it won't be—although I could still learn about his beliefs from it.

Semantic activity and classification play a large and central role in our lives. They are at the heart of any cooperative activities, even activities whose goal is not the communication of information, and have a semantical aspect. A kiss, for example, may seem a rather non-semantical activity. But

a successful kiss will be the product of semantic classification, for it will be the result of a mutual assessment of the content of the participants' desires. And it may itself be, in part, a semantical activity, a way of communicating the presence (or absence) of further desires.

The flow of content, semantic classification, and semantic activity are as natural and familiar to us as breathing or eating or as the fact that ripe apples fall to the ground rather than rising to the heavens. With each of these latter phenomena, scientific investigation and understanding began by looking at the familiar and natural in a new way, as full of contingency, complexity, and even mystery. We need to capture this same sense of wonderment in the semantic aspects of our life.

For the purposes of instilling such wonderment, the metaphor of flow can be quite misleading. Information is not a substance that gets poured from one vessel—say the vet's head—into another—say Jonny's.

Content properties are in fact very abstract properties of complex events and processes. The "flow of information" involves preservation of these abstract properties in virtue of causal interactions at more concrete levels of light, retina, nerves, mouth, ear, finger, and keyboard. I shall call the events and processes that have content *properties representing situations*. These are situations that somehow point beyond themselves to how things stand elsewhere. The x-ray represents Jackie's leg as broken; Fodor's state of mind represents me as skating on thin (but interesting) philosophical ice. When we are engaged in semantic classification and activity—and things are working well—the representing situations are almost transparent. When I read the newspaper, my mind doesn't pause on the fonts or sentence structure chosen, but goes straight to the situation in Iraq. While a novice at x-rays may ponder the exhibited patterns with great care before coming to a diagnosis, the mind of our experienced vet pauses only momentarily before pronouncing on Jackie's leg. Because the whole thrust of our normal interaction with representing situations is beyond them to what they tell us about, it is unnatural to reverse things as I am doing and focus on what their content requires of the representing situation. But this is what we must do to understand the world of information we live within. So, rather than using representing situations as a way of learning about the world beyond them, the current perspective moves in the opposite direction. A content property is a way of classifying a representing situation, in terms of what it represents. For example,

shows that Jackie has a broken leg

classifies the situation of the x-ray exhibiting certain patterns of exposure;

believes that Jackie has a broken leg

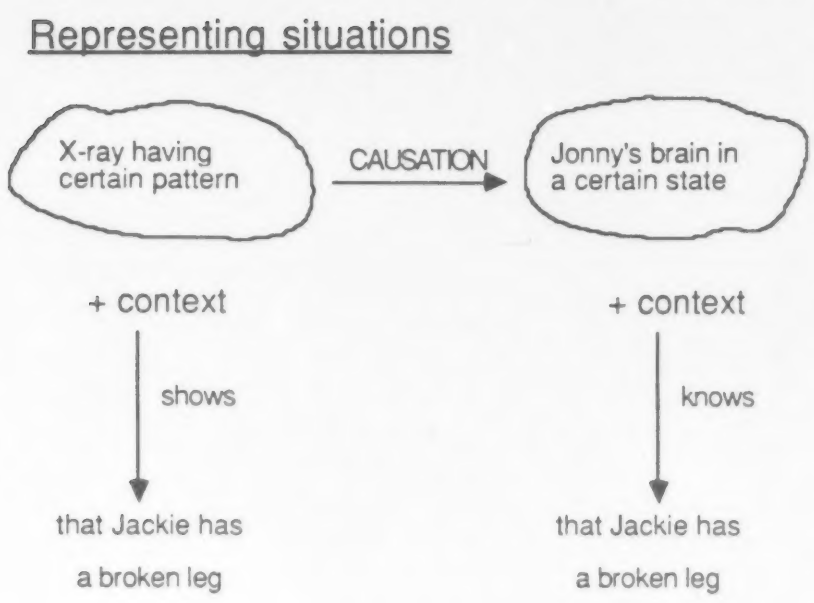


Figure 2.

classifies the state of the vet's brain or mind:

says that Jackie has a broken leg

classifies a linguistic event in terms of its content:

stores the information that Jackie has a broken leg

classifies the state of Jonny's Macintosh, after he goes home and enters the day's sad events into his diary.

The flow of informational content, then, is a very complicated process, whereby a causal sequence of events, involving very different sorts of structures and processes, exhibit, closely related properties—related by having the same content.

The flow of information, thus conceived, is a special case of the more general case of the preservation, transformation, and enhancement of content.

When Jonny learns that Jackie has a broken leg, the information he received from the vet is enhanced by what he already knows. He understands the implications of Jackie's broken leg: that he will not be able to take her for a walk or play fetch with her for a while, that she will be uncomfortable, that the vet will present his parents with a bill. While we rely on each other to preserve and transmit information, it is a sad fact that some reliably distort rather than preserve information. Some people and newspapers make a habit of transforming information transmitted to them into misinformation that they transmit to others, and my TV set regularly turns information into garbage. When I say "flow of information" or "flow of content," I shall usually mean not just the preservation of information, but also such enhancement and transformation.

In our sequence of representing situations, we have two sorts. That the x-ray shows that Jackie has a broken leg does not imply any intentional activity on its part, but only that it contains information that the vet recognizes. We would not ascribe *intelligence* to the x-ray just because we attribute the content to it. But what we say about the vet and Jonny imply semantic activity and intelligence on their parts. When the vet sees that Jackie has a broken leg, he classifies the x-ray as showing that she does. When Jonny hears that Jackie has a broken leg, he is classifying the vet as saying that she does. If Jonny says to his mother that Jackie has a broken leg, he engages in linguistic activi intended to have a certain content. Properties of the second sort, that imply semantical classification and activity, I shall call "content attitudes" or simply "attitudes."

We might explain certain things that intelligent agents do with the x-ray in terms of what it shows. The fact that it shows that Jackie has a broken leg explains, for example, the vet's putting the x-ray in the file marked "Jackie" and not in the one marked "Molly." But apart from this, the content properties of an x-ray didn't explain much about what happens to it. The x-ray will decay or burn as fast whether it shows that Jackie has a broken leg or that Molly does or shows nothing of interest at all.

Attitudes, on the other hand, explain a lot. Jonny is sad *because* he believes that Jackie has a broken leg. It is because he believes this that he says what he does to his mother and father. Because his father hears that Jackie has a broken leg, he checks his bank balance. And so forth. We are very much *controlled* by our attitudes.

The fact that we are controlled by our attitudes is presumably due to the advantages such control offers in surviving in a complex environment. But there now exist things, namely computers, which were designed to be subject to such control, to exhibit intelligence.

One may resist attributing intelligence and the properties that imply it, such as knowledge and belief, to computers. Perhaps one thinks that they are not conscious and that consciousness is implied by "intelligence." Or perhaps one thinks that terms such as "believes" and "knows" stand for properties that only organisms, and not entities built of silicon, can exhibit. Or perhaps one feels that, although silicon-based entities might know and believe and think, the manner in which presently existing computers work is just too different from any conceivably correct account of how humans do for such attributions to be appropriate.

There are important philosophical and empirical issues behind such worries. They should not be dismissed. But they can be, so to speak, tabled. There is no doubt that a battery of concepts that imply semantical classification and activity are appropriate to computers. Jonny's Macintosh not only stores the information that Jackie has a broken leg, it can also compute, on the basis of dozens of separate pieces of information, that Jonny's father owes the IRS a lot of money. Even if we should preface each attribution to computers borrowed form language about humans with "artificial," the fact remains that "artificially remembers that" and "artificially believes that" would themselves stand for attitudes, properties computers have because they were designed to pick up, combine, enhance and transform content, and generally to play an active role in the flow of information.

An analogy might be helpful. Birds fly and airplanes fly. There are principles governing flight that govern both and are thus relevant to both aeronautical engineering and ornithology. The fact that airplanes fly differently than birds do is not unimportant, but it shouldn't blind us to these general principles of flight. Just as we need not believe that birds are really airplanes to recognize the subject of flight, we need not accept the "computational hypothesis" that human minds are biological digital computers, to recognize the subject of information and intelligence.

Thus we are led to recognize a single rather abstract subject matter, the study of information and intelligence, that comprehends parts of a number of disciplines.

We can understand their relationships in terms of Figure 2. We need to understand the nature and structure of the representing situation. In the case of a linguistic event, this would involve phonemics, phonology, morphology, and syntax. We need to understand the relation of this structured situation to its content. In the case of a linguistic event this involves natural language semantics, a field cultivated by linguists, philosophers, and logicians. The causal connections between the representing situations need to be studied not only at the physical level but also at the abstract content level, for it is the principles at the latter level that we use in predicting

and explaining each other's behavior. So, for example, the interaction between Jonny and his dad when he tells him of Jackie's problem requires the study of Jonny's tongue and his Dad's ear, but also of how intention and belief control the content of utterances and how knowledge and expectation govern their interpretation.

Thus research into Information, Computation, and Cognition calls upon (at least) the following (already overlapping) disciplines:

- perceptual psychology, to analyze the pick-up of information by organisms;
- cognitive psychology, to analyze the processing of information by organisms;
- psycholinguistics, to analyze the links between cognition and language use;
- computer science and AI to develop and articulate the principles by which machines can be designed to engage in semantical classification and activity;
- linguistics, to analyze the structure of natural languages, the most successful means we have for communicating complex information;
- philosophy and logic, to develop mathematically rigorous and conceptually clear theories of informational content, attitudes and action, that will enable us to see how representing situations can be classified and controlled by content properties.

Research into information and intelligence is thus inherently interdisciplinary. It is unavoidably chaotic? I believe it is not. On the one hand, linguistics, psychology and computer science are well developed disciplines: we know a lot about the structure of language and a lot about how people and computers work. On the other hand, general conceptions of information and action are emerging from the work of logicians and philosophers and computer scientists engaged in artificial intelligence that promise to provide the unifying conceptions needed by a science of information and intelligence. The basis for these promising developments is a renaissance in logic, which forms my next topic.

Semantics

What is information content?

Luckily, we need not await a definitive philosophical answer to this question to begin developing and using theories of content. To take two rather close analogies, first, the development of number theory—an abstract theory that underlies our application of the same abstract objects, numbers, to diverse phenomena through systems of measurement—did not need to await a philosophical understanding of what numbers are. And, second, the

theory of probability has not awaited agreement among philosophers, or probability theorists, about whether probabilities are measures of relative frequency, subjective certainty, or something else. I don't want to demean the efforts of philosophers, of course. I strongly suspect philosophical worries are a necessary condition of the development of such rich scientific tools. But philosophical agreement is not.

A useful theory of information content need not give us the ultimate nature of information. But it must enable us to represent precisely the crucial features of content. There are three criteria for such a theory I shall dwell on here. First, it must give us a satisfactory notion of identity and diversity of content. Second, it must give us an account of the *direct classification* of situations by content, that is, an account of which situations make a given content true. Third, it must give us an account of the *semantic classification* of situations by content, how objects can have what we have called "content properties," including attitudes.

Let us illustrate some of these considerations.

Suppose that ten years ago I wrote the following in my diary:

1. I am from Nebraska;
2. I have brown hair;
3. I am six feet tall.

Let's start with a really bad account of content. We might measure the *amount* of information by the number of words or characters. By the first measure, 1 and 2 have the same amount of information. By the second, 1 and 3 do. Now clearly, amount of information, conceived in either way, is not the same as information content. For there is no reasonable way of looking at it that gives any pair of 1-3 the same content. A theory that equated amount of information with informational content would just be bad.

No one has ever held this theory of content, as far as I know. But I do have an ulterior motive in mentioning it. There is one aspect of the theory of information that is well developed and has been tremendously useful. This theory is often called "the theory of information" or "information theory" or "Shannon's theory of information." So, having heard of it, one might wonder what more we would need. The answer is, we need a great deal more if we are to deal with content. The theory of quantity of information is not a theory of informational content.

Here is another, more promising theory. Why not just take the content to be the linguistic expression used? This would give us the result that no pair of 1-3 have the same content, since no pair use the same expression. This seems a pleasant idea, for we already have a discipline, linguistics, that is charged with telling us about expressions (as well as a lot of other things).

Suppose, however, that I now pick up my diary, ten years later. If I were now to write in it:

2a. I have brown hair;
3a. I am six feet tall.

these statement would not have the same content as my earlier ones. They do not contain or convey the same information. The stress of thinking about information (and raising funds for this activity) have left their mark; my hair is grey and I have shriveled up by at least an inch. The same expression used in different situations can carry different information.

An example from the first section makes the same point. When I said "I am from Nebraska," I said something true that has the same content as what you said when you said to me, "You are from Nebraska." But if you had used the same sentence I had and said, "I am from Nebraska," you would have said something different, and, most likely, something false.

There are a couple of points these examples illustrate. The first is that if we restrict ourselves to too narrow a class of representing situations, the problem of content may seem more trivial than it is. If we restrict ourselves to situations in which one person says certain things at one time, then the identification of content with expression may seem pretty harmless. When we cast our net wider the identification inevitably fails. To take a much different sort of example, suppose a computer scientist is tempted to take the content of a program to be the internal states of the computer on which the program is compiled. In a sense, this seems right; any other program that gave rise to the very same internal states would be equivalent in content. But when one casts one's net wider, and looks for a notion of content by which programs running of quite different computers with different architectures and operating systems may be compared, this proposal is bankrupt. Note that our ordinary notion of content is extremely versatile; that is the point made by the flow of information examples. We somehow understand the very same content as being what is shown by the x-ray, noticed by the vet, and later stored in Jonny's Macintosh.

The second point has to do with the dependence of content on context and the consequent dependence of recognition of content on knowledge of (or attunement to) context.

Because I wrote "I have brown hair" on a certain day ten years ago—and looked at myself in a mirror to check before writing—the entry in my diary carries the information that ten years ago I had brown hair. But to recognize its content and gain information from looking in my diary, one has not only to recognize the characteristics of the sentence used, but also know the context in which it was written. In general, the content of a representing situation depends on its context, and not just its "internal" characteristics.

The x-ray provides another example. It showed that Jackie had a broken leg. But an x-ray with the very same patterns of light and dark areas could have been produced by x-raying a different broken-legged dog. It is *of* Jackie because Jackie played a certain role in the production of the x-ray. The informational content depends on the context as well as the internal characteristics of the x-ray. And, if the vet was disorganized and didn't label his x-rays as they were taken, he could not recognize the content of his own x-ray.

I want to shift to another type of example, which will emphasize this fact about recognition of context and content, and also suggest how important these matters are when it comes to getting computers that can communicate with language in something like the way we do.

I have borrowed two examples from Barbara Grosz's work on discourse.

In the first, we have the pronoun "it" occurring thirty minutes after its antecedent. Here is the situation. An expert is instructing an apprentice in how to assemble an air compressor. First the apprentice asks, "What do I do first?" and the Instructor replies "Assemble the air compressor." Thirty minutes later, with no further mention of the air compressor, the expert says "Now plug *it* in and see if *it* works." The information that the apprentice picks up—as do any of us examining even this bare outline of what was said—is that the air compressor is to be plugged in. But the sentence in question does not by itself contain this information. "It" could be taken to stand for any number of things that have been previously mentioned and are of the right gender: the screwdriver, the wrench, the toolbox, the flywheel. All these were mentioned during the thirty minutes. A computer doing a backwards search for antecedents that fit would give up long before it got back to "the air compressor." And yet we recognize immediately that this is the correct antecedent.

Here is another example that makes the same point:

LIBERATED HUSBAND Jane left the groceries for me to put away
stop that you kids!
and I put them away after she left.

We take the antecedent of "them" to be "the groceries" and don't assume he put away the kids right after his wife left.

In each of these cases, what enables us to ascribe the right content to (i.e., to get the right information from) the utterance is our knowledge of the context and its relevance. The crucial point is that a great deal of what we have to know is not in any reasonable sense linguistic. In the second example, we know that putting kids away is a much more serious business

than putting away groceries. In the first example, we have an idea about the structure of the activity of putting together a piece of machinery, we know that air compressors typically have plugs while tool-boxes do not, and we know what the shared goals of the expert and apprentice are likely to be. In taking the content of "Plug it in" to be that the air compressor is to be plugged in, we assume a great deal about the beliefs and intentions of the speaker; these assumptions range over all sorts of disciplines. The interdisciplinary entanglements of a subject like discourse should be clear.

A next candidate for content might be that content is just the "ideas" in the mind of the person who utters the linguistic expression. After all, the linguistic expression itself would not convey information if it were not backed by the proper beliefs. Isn't it the ideas, or more generally the cognitive states of the speaker, that constitute the informational content of the utterance? We *might* suppose these ideas to be somewhat more explicit than the sentences used to express them, so that the problems of undetermination of content by expression would not be applicable to this suggestion.

There is clearly much that is right about this. Linguistic events by and large do not carry information except insofar as they are the product of mental states and activities that carry that information. But there are a number of problems, of which I will mention only the most important.

In speaking of ideas and thoughts we are *already* employing content to classify things. We identify even our own ideas and thoughts in terms of objects and situations outside of the brain:

My idea *of* Washington D.C.

My thought *that* someone is going to have a lot of cherries to pick.

There is plenty of room for our sense of wonder about content properties to get a grip here. Can it really be that we classify our own states of mind, that part of reality we seem most close to, in an indirect fashion, in terms of objects and situations outside of them? That would be surprising—and yet try to state something significant about your own desires and beliefs and their interaction, while classifying them in some other way. This is not an objection to our vocabulary of ideas and thoughts, nor to more well-honed terminologies of cognitive psychology. Rather, it is one more tribute to the power of semantical classification—of dealing with things in terms of their content. But it leaves the problem of content unresolved.

It is tempting to respond that we are classifying our ideas and thoughts not in terms of content, but in terms of the language that would express them; not Washington D.C., but "Washington D.C." gives the content of my idea. But here the circle gets pretty tight. This back and forth in the

history of philosophy and psychology, between interpreting words with ideas and then ideas with words, will be a theme of Jerry Fodor's talk.

I hope to have convinced you that a theory of informational content is not a trivial matter. But is it possible at all? I think the answer is positive; logicians are providing us with ever richer theories of content, increasingly well suited to the needs of a discipline of information and intelligence.

This seems appropriate. Logicians, after all, are paid to tell us about truth and validity. We noted that part of the story of content was that contents directly classify situations, as those that make them true or false. If logic can get this much of the story straight, it seems like the role of content in semantic classification should also be illuminated. Things are not quite this simple, however.

Here is an (overly simple) example of a treatment of truth, using some ideas from logic:

Language L

Predicates:	F, G
Terms:	a, b
Logical symbols:	*not*, *and*, *or*
Syntax:	If $\mathcal{F}$ is a predicate and α is a term,
	$\mathcal{F}\alpha$ is a sentence
	If φ and φ' are sentences, then
	not (φ), $(\varphi$ *and* $\varphi')$ and $(\varphi$ *or* $\varphi')$.

Semantics:

Dictionary:	F stands for the set of things with broken legs
	G Stands for the set of things with tails
	a stands for Jackie
	b stands for Molly
Truth:	$\mathcal{F}\alpha$ is true if α stands for a member of what $\mathcal{F}$ stands for.
	not (φ) is true if φ is not true
	$(\varphi$ *and* $\varphi')$ is true if φ and φ' are both true
	$(\varphi$ *or* $\varphi')$ is true if at least one of φ and φ' is true.
Validity:	φ is valid if φ is true no matter what appropriate dictionary is substituted for the right one.
Examples:	Fa, Ga, and Gb are all true; $(Fa$ *or not* $(Fa))$ is valid

This is extremely simple; but, from ideas not much more complex than these, logicians have built a remarkable edifice of results about truth and validity, a solid part of mathematics.

The construction of such theories is called semantics. Does logic, then, provide us with what we need for a theory of information? We need to

focus on two ideas exhibited here, a promising one and another that for our purposes is less promising.

First, consider the idea of assigning semantic values (truth values) to complex expressions (sentences) on the basis of primitive assignments (the dictionary) to basic expressions. This seems like a promising idea for any theory of ascription of content. Analyze the structures of the representational events in question; assign objects to these aspects; then systematically assign contents to the possible structures on the basis of those assignments.

But, secondly, there is something distinctly unpromising about the choice of semantic values here. Truth values are simply not good candidates for content when we are interested, as we are, in the indirect or semantical classificatory powers of content. For *Fa* and *Ga* and *Gb* will all have the same content—the truth value true—and so

1. believes *Fa*
2. believes *Fb*
3. believes *Ga*

will all have exactly the same indirect classificatory powers. But where Jonny's being in the state classified by 1 led him to be sad, being in the states classified by 2 and 3 would not have. Truth values cannot serve as content. They are too "coarse grained" for a theory of semantical classification.

We might try to improve things by moving to the notion of truth relative to a dictionary. *Fa* and *Fb* will be true relative to different dictionaries, for *Fa* will be true relative to any dictionary that makes *a* stand for a member of what it makes *F* stand for, and such a dictionary might make *b* stand for something not a member of this set.

But this idea is really rather odd. It is the dictionary that gives us the connection between the language and the world. On the proposal, a dictionary that assigns to *a* the number 1 and to *F* the set of odd numbers will be as much a part of the content of *Fa* as one that assigns Jackie to *a* *and* to *F* the set of broken-legged things. How could such a notion of content capture that which allows us to get at both what the x-ray shows and Jonny believes? The x-ray is *of* Jackie and Jonny's belief is *about* Jackie. But Jackie plays no bigger role in the set of dictionaries that make *Fa* true than, say, Molly, who will be what *a* stands for in just as many of these dictionaries as Jackie.

Is this the best the logician can do for us? For a long time, it seemed as if it might be. Semantical systems with truth in a dictionary (or "model") as the primary notion of content have proved so successful in the logical analysis of mathematics that such an approach almost came to be equated with logic itself. The lack of alternatives that might be suited to give us

a theory of content even led some philosophers to despair of making sense out of our semantical classifications and assessments.

The situation has been changing steadily over the past thirty years, in a movement that continues to pick up steam, which I call the Renaissance of Logic. Recall that theoretical activity was not dead during the Middle Ages, it was just narrowly focussed on theology. The renaissance contributed a wider conception of what was worth thinking carefully about. Once the empirical world that humans inhabit became a respectable object of study, new methods, both humanistic and scientific, quickly followed. Similarly, logic was never still, but just for a long time rather single-mindedly devoted to understanding languages designed for the eternally perfect realms of mathematics. The Renaissance in Logic consists of an expanded conception of what can and should be understood using logical methods.

Two founders of modern logic, Gottlob Frege and Bertrand Russell, were actually quite concerned about indirect classification, and tried hard to develop systems that took attitude verbs like "believes" seriously. Their efforts to do so were not successful, however, while the "extensional" part of their logic was fabulously so. The latter was put on firm conceptual and mathematical foundation by Gödel and Tarski, and the mathematical discipline known as model theory emerged from their work. Interest in a wider set of problems for logic to handle was kept alive by Alonzo Church, who formalised Frege's theory of sense and reference. But a somewhat different approach, stemming from Frege's student Rudolf Carnap, has been more influential.

To get the flavor of Carnap's approach, return to our last seemingly confused idea, that sets of models or dictionaries serve as contents.

A model is a function that assigns extensions to predicates—bits of language—directly. But note that such a direct assignment treats as one step what is really two. Why is Jackie in the extension of "has a broken leg?" First, because "has a broken leg" has the meaning it does. If it meant what "has a Ph.D." in fact means, she wouldn't belong in its extension. Second, because in fact her leg is broken; if it were merely sprained, she also wouldn't belong in the extension of "has a broken leg."

In studying the languages of mathematics, models are thought of as alternative dictionaries. Given the non-contingent nature of the facts of mathematics, there is really no need to worry about alternatives of the second kind. Jackie's leg might not have been broken, but there is no possibility of two not being an even number. But in studying languages that concern the empirical world, such contingencies need to be represented.

Carnap saw that models could be interpreted as alternative "possible worlds" as well as alternative dictionaries. We think of the predicates as

having a fixed meaning, and the models as showing how the facts can vary. Conceived as a set of possible worlds, the set of models relative to which *Fa* is true is a much more promising candidate for its content.

Such a theory of content is still not fine-grained enough, however. We saw that the attitudes are characteristically "self-embeddable." Jonny not only believes that Jackie has a broken leg, he believes that the vet believes this; it is because he believes the latter that he believes the former. But Carnap's approach won't work for beliefs like the latter. The contents cannot be more fine-grained than the alternative extensions that can be assigned to the original stock of predicates. But with the original stock, we cannot represent the difference between a world in which the vet believes that Jackie has a broken leg, and one in which he doesn't. We can't solve this by adding a predicate for "believes" to the original stock, for we need sets of models to represent the extension of this predicate, but for each model in the set, we would already need an extension for it. This problem can be finessed in various ingenious ways, but it never goes away.

The renaissance began to gather steam with an idea of Saul Kripke's. He saw that the two different jobs models were used for—as dictionaries and as possible worlds—needed different bits of machinery in the same theory. He took possible worlds to be real objects (leaving to philosophy the job of determining just what the objects were). An *intension* is a function from possible worlds to an extension—a set of individuals. The dictionary now assigns intensions to predicates, rather than extensions. Whether "has a broken leg" means having a broken leg or having a Ph.D. depends on the intension assigned to it. Whether Jackie has a broken leg or not depends on which possible world is actual.

Here is a "possible worlds semantics" for our language L:

- F stands for a function f_F from possible worlds w to the set of things that would have broken legs if w was actual;
- G stands for a function g_G from possible worlds w to the set of things that would have tails if w was actual;
- a stands for Jackie
- b stands for Molly
- The *content* of a sentence $\beta\alpha$ is the set of possible worlds w such that α stands for a member of $f_\beta(w)$.

Notice now that *Fa*, *Ga*, and *Gb* have quite different contents, even though they have the same truth-values. Thus, we begin to have a chance at a theory of information that can account for semantic classification.

Another important development are logics that take context seriously. We have noted content typically depends not only on the internal characteristics of the representing situation, but on the external context as well.

Mathematical utterance seems to be atypical in this respect. If I had written "2 + 2 = 4" in my diary ten years ago and wrote it there again now, the content of the two statements would be the same. Here again, there was a period in which it was assumed that if a phenomenon didn't fit the logic that sufficed for mathematics, it was the phenomenon, not the logic, that was at fault. But now a variety of logical systems are under development that relate content systematically to the context of the representing situation as well as its characteristics.

The basic idea of this development, due to David Kaplan, is to distinguish between the meaning or character of a sentence and the content of a particular use of it. Consider the predicate "is kissing me." As we saw, a dictionary shouldn't simply assign an extension to this, for the extension depends on the facts—which possible world is actual—as well as the meaning. But the dictionary shouldn't simply assign an intension either. The set of people in question when I say "is kissing me" may be quite different from the set in question when you utter these same words, even when we are considering the *same* possible world. Since in the actual world I am busy talking and you are busy listening, it happens to be the same set—the null set. But if you had stayed home, the sets might be different. So the dictionary should really just assign what Kaplan calls a *character*, a function *from* a context (with you, I, or someone else as the speaker) to an intension.

The developments just sketched were all due to logicians who worked with artificial languages. Another tradition in logic that took a long time to fade was the idea that only such artificial languages are susceptible to precise semantic analysis. This tradition died suddenly, with the work of Richard Montague. Montague, who was also involved in the development of possible worlds semantics and treatments of context, showed how semantic analysis could be extended to natural languages. More recently, Dana Scott has shown how programming languages can be given a precise semantical analysis.

I have tried to give a hint of a tremendous amount of work, which has revived the hope of the founders of modern logic, that logic could serve as the means to bring precise mathematical analysis to a host of disciplines that deal with content. Those hopes have not yet been achieved, but they are within sight.

Thus, a theory of information that can provide the mathematical underpinnings for cognitive semantics is neither trivial nor impossible. A number of promising approaches are being developed, and these developments in logic are having ever more profound influence on the way that linguists, computer scientists, psychologists, and others approach their subject matters.

What's Happening

The study of Information, Computation, and Cognition is going on full blast. Interdisciplinary groups—like those at M.I.T., Texas, UMass, Pittsburgh, Penn, Stanford, and elsewhere represented here—have sprung up around the country. This research benefitted enormously from NSF grants given to individual researchers in the involved disciplines and by the avowedly interdisciplinary programs of the Sloan and System Development Foundations, which have sponsored the formation of centers at various universities around the nation.

These scientists, collaborations, and centers are a national treasure. There is one aspect of this I would like to briefly emphasize, at the risk of being too practical for the present audience.

Stan Peters has remarked on the development of AI in America, from the Age of Hubris, to the Age of Appreciating Other Disciplines, and now into the Age of Semantics. This is by no means a world-wide phenomena. It is still possible for computer scientists to believe that only more powerful CPU's and larger memories stand between us and the promise of artificial intelligence.

The AI-ers in America, or a significant subset of them, know better. They have recognized a **theory gap.** Just as calculators could not have been built without arithmetic, so, too, programs that approach human capacities in thought and action, even in limited domains, cannot be built until we have better theories of informational systems to conceptualize what needs to be done. The appreciation of this by the generation of Artificial Intelligentsia represented in this room is a key reason we have the high degree of interest in cognition and information that we do. To drive this point home, let me return a moment to the discourse examples mentioned above.

Any human immediately recognized the antecedents of the pronouns in these examples. Our ability to do so, however, is not a simple matter of understanding facts about the language used. Think for a moment about the knowledge you brought to bear in the first example. You knew how an apprentice and an expert would interact. You had some idea of the activity of putting things together. You knew the likely content of the intentions and beliefs of someone who would say, at a point in the activity, "Plug it in." You knew that air compressors, but not wrenches and screwdrivers, have plugs. If the whole discourse were placed before you, or if you participated in it, you would easily sense the structure, the shifts of focus and intention that occur within it. Even for the relatively limited project of building a program that in this one area, teaching someone to assemble an air compressor, one needs some understanding of how such capacities might work. We will not have such an understanding without a concerted interdis-

ciplinary effort involving researchers in areas from phonology and computer vision to the representation of human knowledge to abstract logic. No guarantees can be made about when the prospects of AI will bear real fruit if support is forthcoming for its scientific underpinnings; but this much can be guaranteed: given the interdisciplinary nature of the needed work, the size and the complexity of the problems, without such support, the dreams of AI will remain dreams for a long time.

A Presentation to the National Science Foundation Workshop on Information and Representation

Jerry A. Fodor

Most of what I'm about to say concerns the status and prospects of empirical research—narrowly in cognitive psychology and, more broadly, in cognitive science. But, like the previous speaker, I am a philosopher by training and philosophers are prey to a variety of worries that are specific to their profession. There are, for example, those of us who fear that there is something fundamentally unsound about tables and chairs; that perhaps we shall run out of prime numbers or afterimages; that our bodies aren't really there or that our minds aren't; stuff like that. I am aware that, to our colleagues in other disciplines and to our spouses, these anxieties often seem extravagant. But when they're added to the usual concerns about getting tenure, I can assure you that they constitute a heavy load to bear.

The reason I am bothering you with these markedly uninterdisciplinary reflections is this: Every now and then some merely philosophical worry turns out to be real. It thereby ceases to be *merely* philosophical and presents itself as a matter of general concern. I think that this is now happening at the core of cognitive science and that it has implications for the organization and administration of cognitive science research. Problems about the representational character of cognition have, for centuries, been the philosopher's proprietary turf if only because nobody else seemed to want them. But they are now becoming pressing in a number of empirical disciplines, and their solution is increasingly the precondition of further progress in these sciences. Moreover, I sense a rather widespread—if implicit—consensus that something of this sort is true. I propose, in the next less than half an hour, to try to make this consensus more articulate.

About two hundred years ago, philosophers began to be depressed by the following thought: that problems about the cognitive mind and problems about symbolic representation would have to be solved *together*. This thought is depressing because problems about mind and problems about representation are both very hard, and when you've got one very hard problem to solve the last thing you need is another one. In fact, it's so depressing that the fear of having to face it has made behaviorists out of generations of philosophers. And, of course, psychologists.

Depressing or otherwise, however, it's extremely plausible that the relation between cognition and representation is intrinsic. There are lots of ways of making this point. Here's one: What philosophers call "'intentionality"—the property, to put it briefly, of *aboutness*—bifurcates the world. There are, as far as we know, two *and only two* sorts of things in the whole universe that have it, these being *mental states* and *symbols*. So, suppose, for example, that J.—S.J. Keyser, as it might be—believes that Chaos is up again. (Chaos is a DEC-20 that lives at M.I.T. Like everybody else at M.I.T., Chaos is neurotic and a workaholic.) Then J.'s belief has a certain propositional content; viz., it has the content *that Chaos is up again*. In virtue of having that content, J.'s belief is about something (viz., it's about Chaos's current condition). And in virtue of being about Chaos's condition, J.'s belief is semantically evaluable; viz., J.'s belief is true on the off-chance that Chaos is working, and false otherwise.

Much the same sort of thing, however, can be said about *symbols*—forms of words, as it might be. Consider, in particular, the English form of words "Chaos is up again!" It looks as though that sentence—or perhaps that sentence as employed in certain contexts—has a propositional content (it expresses the proposition that Chaos is up again); and it's about something in virtue of having the content that it does; and it's true or false depending on how things are with what it's about, viz., on how things are with Chaos.

Notoriously, it's hard to say all this just right; that, in fact, is one of the things that philosophers worry about. But the way I've just said it should be good enough to make the point that I'm urging. Nothing *remotely like* this story about mental states and symbols applies to any of the rest of the created universe. There's nothing that trees, or prime numbers, or protons, or tables and chairs, or spouses are about; they have no propositional content; and they are not *semantically* evaluable, however useful they may be in other ways. Well, it's hard to believe that this is all just accidental. And if it's not, it's hard to see how whatever theory works for symbols could fail to connect pretty closely with whatever theory works for mind. It looks as though the mental and the symbolic form a *natural theoretical domain* and are both going to have to shelter under much the same umbrella.

So much for a quick glance at why it might seem *plausible* that we need a single theory for cognition and representation; now for a glance at why it might seem sad.

If one looks at early attempts to construct the required covering theory, the situation doesn't seem at all bad. Take somebody like Hume, for example. Here's Hume's story about how you connect the theory of cognition with the theory of representation: Minds, he says, are populated

by a class of particulars called *ideas*. Ideas come in roughly three kinds: sensations, perceptions and thoughts. (I'm not actually being faithful to Hume's terminology, which tends to be imperspicuous.) Now, seen from the psychologist's point of view, the interesting fact about ideas is that they enter into relations of association. Association is a causal relation in virtue of which one idea conjures up another; Hume describes it as the mental analog to gravitational attraction. So, then, once the outside world starts things up by presenting a sensation to the mind, association can keep things going by calling up perceptions and thoughts to keep the sensation company. Hume assumes that the paradigmatic cognitive achievements—specifically, thinking and what modern psychologists sometimes call "perceptual integration"—are instances of this sort of chaining, so Hume's theory of cognitive processes reduces exhaustively to his theory of association. A fact that Hume is pleased with and advertises frequently.

Seen from the semanticist's point of view, however, the interesting thing about ideas is that they are *symbols*. Give or take a little, Hume supposes that ideas are *images*, and that they represent what they resemble. Notice that Hume really does *need* some such semantical assumption because he has to answer the question "How could the associative chaining of ideas constitute *thinking*?" In particular, chains of thought are sequences of mental states; and, as we've seen, *mental states typically have intentional properties; they're about how the world is*. So, if chains of thought are chains of ideas, then ideas must somehow have intentional properties too; ideas too must somehow manage to be about how the world is. So Hume makes this characteristically brilliant move: Ideas have intentional properties *because they are symbols*; the semantically evaluable thus *reduces to* the symbolic; the resemblance theory of representation *explains* the intentionality of cognition.

To summarize, by the end of the 18th Century, more or less, we have the Humean synthesis in place: the connection between the cognitive and the representational is explained by the hypothesis that mental processes are operations defined on mental symbols.

We now skip a couple of hundred years during which nothing of significance transpires. And arrive at The Modern Era.

Here's what happened in The Modern Era: First, the Humean synthesis broke down. Second, its breakdown came to be widely noticed. Third, the fact that the Humean synthesis had broken down—and that no adequate replacement was in view—began to present an impediment to empirical research in a variety of disciplines from, say, linguistics to computer science to AI to cognitive psychology to neuropsychology to the philosophy of mind. The third of these events is patently a cause for interdisciplinary concern.

It constitutes, in my view, something of a scientific crisis but also an unparalleled research opportunity. And it's what brings us here today.

The crisis arises because only half of the Humean synthesis has proved tenable. The idea that mental processes are defined on symbols underlies all the best modern work on cognition: in the current state of the art, it's simply irresistible. But the idea that mental symbols are images now appears to be subject to insuperable objections. So we are in the position of having a representational theory of the mind without having a theory of mental representation. This is not a stable position: something has to be done.

I expect that much of this part of the story is familiar to most of you: but let me try to put it in synoptic form. We are, I think, badly in need of perspective.

There may be mental images, and it may be important that there are some. (I'll return to this presently; it's one of a budget of empirical issues in psychology that our inadequate understanding of representation makes it hard for us to work on). But, however that goes, Hume's hope that we might solve the problem of mental representation at a stroke by identifying the intentional with the imagistic is clearly doomed. For two reasons. First, if you're going to identify thinking with having causal sequences of mental symbols run through your head, then **the vocabulary that these symbols make available must be rich enough to express all the contents of thoughts.** If, for example, thinking that Chaos isn't up is having a negative thought, and if having a negative thought is entertaining a mental symbol that expresses a negative proposition, then **mental symbols must be the sorts of things that can express negation.** It is, however, very hard to express negation with an image (what does a picture of Chaos's not being up look like?) Similarly with hypothetical propositions, and quantified propositions... and so forth. Images just don't have the right logical syntax to do what Hume wanted them to; viz., provide the format in which our thoughts are couched. Try to conjure up an image of "Chaos will possibly be up by next Thursday if J. hasn't forgotten to call the repairman from DEC."

And, second, even if we did think exclusively in images, Hume's idea that the semantics of images reduces to resemblance won't work. A picture can perfectly well resemble—indeed, it can resemble as closely as you like—something that it's not a picture *of*. In consequence, the image theory of thought wouldn't explain the intentionality of thought *even if it were true*. (The Mona Lisa resembles a good copy of the Mona Lisa more than it resembles the Lady who sat for it. But for all that it's a picture of the Lady, not of the copy.) A line of philosophers from Wittgenstein to Nelson Goodman have made a living off this point, and I think that we had better let them have it.

Anyhow, if a language is needed that's rich enough to couch thoughts in—rich enough to provide the arrays of symbols over which cognitive processes are supposed to be defined—we've now got better candidates than images. We've got "formal" languages (from logic) and programming languages (from computer science). For that matter, we've got *natural* languages (for that small and eccentric band of theorists who think that we think in English.) The thing about these sorts of symbol systems is that they provide formal vehicles—logico-syntactical structures—capable of expressing very fancy propositions indeed, symbolic structures in which you get not just negative operators, but modal ones too—to say nothing of variables together with quantifiers to bind them, verbs of propositional attitude with sentences subordinated to them, tense relations of Baroque complexity, and adjectival and adverbial modifiers piled up until the cows come home. The closer you get to natural languages, indeed, the closer you get to representational systems in which one can say whatever one can think. Not surprisingly, of course, natural languages are *for* saying what one thinks.

So what is there to be depressed about? It seems that we actually know some things that Hume did not. On the one hand, we have systems of logico-syntactical objects asymptotically rich enough to provide the vehicles for thought; and, on the other, we have the computer as a model of the mechanical manipulation of these logico-syntactical objects. We can therefore replace Hume's story about images associatively interconnected with an updated story about logico-syntactic forms subsumed by computational operations. We thus have a theory of cognitive processes recognizably in Hume's spirit, but vastly richer in the kinds of mental lives it can describe.

Take linguistics, for example. Here is how we do linguistics: we specify—axiomatically—a formal language (a "level of linguistic description," to use the technical jargon.) This formal language includes a (typically infinite) array of well-formed formulas, together with their intended semantic interpretations. (So, for example, the language might consist of an infinite set of labelled tree structures, specified in a vocabulary that includes the symbols "NP," "VP"... etc. together with salient geometrical relations like domination, concatenation, and so forth.) The intended interpretation of this language is specified by saying such things as that the symbol "NP" denotes the property of being a nounphrase, the symbol "VP" designates the property of being a verb phrase... etc., and that if, in a tree structure, a node A dominates a node B, then the semantic interpretation of this geometrical configuration is that B is a constituent of type A. (I am putting this rather loosely, but I take it that the idea is familiar.)

If all this goes right, it gives us a domain of logico-syntactic objects which, under their intended interpretations, are representations of, say, the

sentences of English and over which we can specify a population of mental—specifically, psycholinguistic—processes. So we can, and do, identify parsing with the construction of such trees from input acoustic waveforms. Speech production is conceptualized as the integration of tree structures and their transformation into patterns of phonetic instructions to the articulatory apparatus. And language learning, at the syntactic level, turns out to be the internalization of axioms which enumerate systems of tree structures appropriate to structurally describe the sentences of the language being learned. (We're even getting some mathematics to go with this picture; for a review, see Osherson and Weinstein, forthcoming.)

So far, this doesn't sound much like a crisis. No doubt there's disagreement on lots of details (and on lots of nondetails, for that matter). But this sort of picture has done good work for us recently across a whole spectrum of disciplines. Even those who are pessimistic about how far it can ultimately be pushed don't expect to have to give it up tomorrow.

So then what's the problem? The problem is that, when we construct these sorts of computational theories in psychology, linguistics and AI, we are working with notions that we don't really understand. And this chicken is now coming home to roost; what we don't understand about the foundations of computation theory is starting to get in the way of our empirical research.

To take one case—not a small one: It's intrinsic to this sort of theorizing that one thinks of the objects that get manipulated by mental processes as "syntactical" or "formal" and that one thinks of the mental processes themselves as "computational." But we don't really know what that means (just as Hume didn't, really, have any *general* account of what it is for something to be an image or a process of association). Battles about this keep breaking into the press. For example, are the operations that "transducers" perform in perception supposed to count as computations within the meaning of the act? If not, why not? If so, what *wouldn't* count as computation? A lot of the recent sniping back and forth between "information flow" psychologists and "ecological realists" of the Gibsonian persuasion has turned on this sort of issue. It doesn't look to be fully resolvable short of a theory of computation broad enough to say just what the presumed analogy between minds and computers is actually supposed to be. And yet, as many of you will recognize, how you jump on this distinction determines your research priorities and tactics. You approach the experimental investigation of perception differently if you're thinking of organisms as transducers than you do if you're thinking of organisms as inference machines.

Or consider another, closely related point. I said that as the languages available for use as systems of mental representation get closer in their ex-

pressive power to natural languages, they also get closer to being plausible as vehicles for thought. But this has a sinister side: As formal languages get more and more natural **we more and more lose our grip on their semantics.** There is no worked out (to say nothing of workable) specification of the "intended interpretations" of the well-formed formulas of English. And a lot of people—especially around Stanford—are worrying that the reasons that natural language is so good at expressing thought are *essentially* connected with ways in which natural languages *differ* from the sorts of formalisms traditionally used in mathematics and computer science, the latter being the only sorts of rich symbolic systems about whose semantics we now have much of a clue. Indexicality has loomed large, of late, in discussions on both Coasts. I know just enough about this to be convinced that—if that's where the problem lies—then it is a **very hard problem**.

But there's worse; it's not just that we don't know how to specify the intended interpretations of formalisms that look rich enough to serve as "languages of thought"; it's also that we don't know what it is for something to *be* the "intended" interpretation of such a formalism. This is where it really hurts that Hume was wrong about resemblance. For whereas Hume could say "what it is for a thought to be about a chair is for the thought to resemble a chair," we can't say that and *we don't know what to say instead.* There is an intuition that thoughts are about the world in virtue of the way that the thinker is causally connected to the world; that—to put it about as misleadingly as possible—semantics somehow reduces to robotics. But nobody knows how to cash this intuition, and the problems look formidable. The discussion of these issues in the cognitive psychology literature thus far is largely a disaster area, obscured by a pervasive confusion between semantics and logical syntax, and littered with the corpses of dead theories.

This is, to put it mildly, no joke; to see that it is no joke, bear in mind that saying that thoughts have semantical properties —that they are about the world—is just a way of saying that thoughts contain "information" about the world. The *fundamental* idea of information flow psychology is that organisms extract information from their environments; that their mental processes elaborate this information; and that their behavior is determined by the information that mental processing makes available. It is precisely this processing of information that computational psychology is supposed to explain and artificial intelligence is supposed to simulate. *And we don't know what any of this means because we don't know what information is.*

How much of this is just philosophy? What's important about the current crisis of theory is that its effects are now felt broadly throughout the information sciences. Here's an example: I've already mentioned the prob-

lem of imagistic thought. Like the work on language, and the work on early vision, the study of imagery is one of the success stories of cognitive science. A case has been made that the mechanisms of mental imagery overlap with those of visual perception. If this is true, we now have a foot in a door that we badly want to open: the issue of how representation in perception connects with representation in thought. In particular, given the experimental work by Shepard, Kosslyn, Fink, Pinker and others, it's not unreasonable to suppose that some mental representation is iconic; and it's also not unreasonable to suppose that the iconicity of some cognitive representation may be important to the success of some sorts of problem solving (see Johnson-Laird; but also Lance Rips).

That's all fine, but then the trouble starts: when you get past the resemblance theory of reference (which, I take it, nobody now believes) what, precisely, does the claim that mental representation is—or, for that matter, isn't—iconic amount to? How, precisely, are the experiments supposed to bear on that claim (and, if the currently available experiments don't bear, what experiments can we do that would?) What about John Anderson's surprising suggestion that you can't *even in principle* have empirical evidence that decides this sort of question? Suppose I want to build a machine that's smart because it manipulates (not mental sentences but) mental models? How do I go about even *thinking* about building such a machine if I don't know either what a mental sentence or a mental model is supposed to be?

Or take another case fresh from the recent empirical literature. There's some fascinating work by Neil Cohen that's supposed to exhibit a pathological dissociation of knowledge of skills from knowledge of facts. So, for example, Cohen reports the following sort of results. If you take normal subjects and train them, over a series of sessions, on a hard task like reading backward text, you find two partially independent learning effects: on the one hand, the subjects get progressively better at the task, whether their stimulus materials be new or old items; but also, there's a differential facilitation of materials previously seen. Thus, if you've had a number of sessions of learning to read backwards and I give you the stimulus "sdrawkcab," you will, of course, do better than a subject who has had no previous training in the task. But also, if you've seen "sdrawkcab" on earlier trials, you'll do better on that item than you would on matched items that you hadn't seen before. So it looks like there are two learning effects: one that's item specific and one that's not. This isn't, in itself, very surprising; but what *is* striking is the existence of brain damaged populations where these effects dissociate; these subjects exhibit the normal improvement of the skill over practice trials, but *not the normal effect of previous exposure to specific items*. It's as though the subject could learn *how* to read backwards,

but can't learn *that* "backwards" spelled backwards looks like "sdrawkcab." These are, interestingly enough, the same sorts of subjects who, though they exhibit serious short-term memory deficits for names, faces, events and the like, are apparently able to perform quite normally on learning to solve the Tower of Hanoi problem. In the extreme case, their learning of this problem is perfectly normal, but they have no knowledge of having learned it. With practice, the subject achieves essentially error-free performance on the task; but if queried, he'll tell you that he has never seen the puzzle before, that he is behaving at random, and that his solutions are accidental.

I really *do* think that's fascinating; but now, unfortunately, the argument comes to a screeching halt for want of some foundational theory. For, how shall we understand Cohen's discovery in computational terms? Well, one possibility—to which, I believe, Cohen is himself attracted—is that the pathological population can learn "procedural" but not "declarative" information. Fine, but what does *that* distinction come to? If you ask a computer type, he's likely to say that it's a *syntactic* distinction, analogous to the difference between imperative and declarative *sentences*. But it's hard to see how that could be the heart of the matter since, clearly, you can encode what's intuitively declarative information in syntactically imperative form. (Indeed, if you're into production systems, you encode *everything* in imperative form, leaving yourself with no procedural/declarative distinction at all at the level of syntax). Or, if you ask a philosopher about Cohen's data, he'll perhaps remind you of Gilbert Ryle's distinction between "knowing how" and "knowing that." But if you ask him to tell you what *that* distinction comes to, I promise that he won't be able to; not, at least, if what you're wanting is the kind of answer that can be formalized to provide a domain for an algorithm. Which is presumably exactly what computational psychologists *do* want.

I hope I've made it clear that the problem isn't just that we don't have a good theoretical vocabulary to report our data in; it's that the best theory of the cognitive mind going, the representational/computational theory, is based on a group of notions—information, aboutness, syntax, representation, computation, procedure, data structure, and so forth—that are largely unarticulated. In consequence, the computational story about the mind isn't doing for us the sort of jobs that one's best theory is supposed to do: guiding research, rationalizing the investigative undertaking, and making the empirical data cohere. Cohen has, as things now stand, these lovely findings; and he has the computer metaphor. His problem is that the metaphor, insofar as it is *just* a metaphor, doesn't engage the data. What he needs to make sense of his empirical findings is not a metaphor but a worked out notion of computation. He's not the only one.

As John Perry rightly remarked, it doesn't matter if one's best science is philosophically disreputable. On the contrary, that's often a sign of progress since developing empirical theories are always more or less inarticulate. And, anyhow, it's amazing how much disrepute philosophers can learn to live with. But it *does* matter when your best theories fail to organize your best research. That's a crisis; and that's what we've got.

So much for the current causes of discontent. A couple of words on the bright side, and a couple of administrative remarks, and I'll be done.

First, I don't mean to begin to suggest that the empirical work on understanding intelligence—artificial and/or natural—has got to down tools and wait upon the solution of the foundational problems about computation and representation. Take a case from my own interests. I do experimental psychology whenever Merrill Garrett lets me borrow the key to the laboratory, and when I do experimental psychology I care a lot about questions of "mental architecture"; questions that concern the restrictions on information flow between cognitive systems. This is the area in which the so-called "modularity thesis" operates. This research, like most of cognitive science, is relentlessly interdisciplinary; you can't, for example, ask how much contextual information is available to parsing unless you have some idea of what parsing is; hence some idea of how the recent linguistics and psycholinguistics is shaping up. But of course you *can* ask it without having a general account of the nature of information. Such an account would certainly be useful in theory, but—thank Heaven—it's not an absolute prerequisite to practice. It is important to realize, however, that doing one's science this way is like living on credit cards. Comes the end of the month, somebody has to pay the bill. Sooner or later we're going to have to say what it is that flows when information does.

Or, to take another example, it is very much on the long-range agenda of modularity theory to understand the interfacing of cognitive mechanisms; which means that we must eventually confront what is, in my view, one of the great mysteries of human and artificial intelligence: the language/vision interface, the problem of how a person is—and how a machine might be—able to *say* what he *sees*, and to use what he sees to interpret and verify the linguistic messages that he hears. This problem is among the great watersheds of cognitive theory. We will not be ready to face it, to ask just how, and just where, and in just what quantities, information can flow across the boundary between language and vision, until we have a far clearer notion than is now available of *what information is* and, in partiuclar, of how the *same* information can survive the transformation from a visual to a linguistic code and back again.

This is, by the way, the long-range problem upon which I believe that the M.I.T. cognitive science research, with its persistent emphases on language, vision, and modularity, will ultimately converge.

Second, though sooner or later we have to have a theory of representation, we perhaps don't need a theory that works for the *general* case. What we need, if we are to understand and create intelligent systems, is insight into the character of those symbolic structures—mental representations—that mediate cognitive processes. It may be that this means a completely general theory of representation (one that works not just for minds but also for pictures, say, and for the harbingers of Spring); but also it may not. Unlike such broken vessels as "semiotics" and "general semantics" we can afford to start out neutral on that issue. The prospects for constructing a theory of mental representation are, to this extent, better than the prospects for constructing a theory of representation at large. This is one reason why it is absolutely essential that the foundational work should proceed in tandem with detailed empirical studies of cognitive and linguistic processes. We come to understand the design specifications for a theory of information as we come to know what representational capacities intelligent informational processing actually does demand. In my view, that is the very most promising direction for research to take in the immediate and foreseeable future.

So here's the Fodor Plan for the development of a cognitive science. One team starts on roughly the semantic end with the problem of understanding information, computation and representation. The other team starts on roughly the psychological end with the problem of understanding mental processes and states. If all goes well, they converge on a theory of mental representation and there is light at both ends of the tunnel. Individual players are encouraged to change teams frequently in order to keep alert and informed. I am, by the way, entirely serious about this proposal. If I had the responsiblity to organize the next couple of decades of research in cognitive science, that is how I would do it.

Third and last of the substantive points: If it's true that the deep questions are getting harder to avoid, it's equally true that the chances for answering them seem to be improving. If we still haven't bettered Hume's idea that mental processes are intentional because they're defined on mental symbols, at least we have a vastly more powerful intellectual technology than he did to work this idea out with. We have learned a lot about symbol systems and their uses since the 18th Century. The trouble is that the available information is scattered through a variety of traditional disciplines. Nobody knows enough—enough linguistics, enough computer science, enough AI, enough psychology, enough recursion theory, enough formal semantics,

enough logic, and enough neuropsychology, for starters—to do the research that now needs to be done. I think this is going to have to be a joint undertaking; a lot of people who are used to working alone are now going to have to learn to work together.

I close with an organizational remark, based on the last ten years' experience in trying to get this sort of interdisciplinary work off the ground. Conferences aren't how to do it. Conferences help with interdisciplinary consciousness raising, but they are not, in and of themselves, vehicles of scientific progress. What's needed, in my view, is the construction of standing research environments that will encourage sustained interaction among specialists with quite different academic backgrounds; and of pedagogical environments that will encourage young scientists to acquire mixes of competences not even contemplated until quite recently. It goes without saying that this cuts across traditional academic affiliations and raises serious administrative problems. For example, it implies a funding structure for cognitive science (singular) that is less rigidly target oriented than NSF funding in the cognitive sciences (plural) has thus far been. Even the organization of peer review is problematic when the object of the exercise is precisely to create new kinds of expertise and bring it to focus upon new kinds of problems. Where, to put it bluntly, do you find the peers?

So, the administrative and the scientific problems are both hard; and, too, it's hard admitting that the problems have maybe started to outstrip one's training. Nevertheless, it seems to me it's time to get the revolution on the road. Two hundred years is too long to be depressed. I think, in fact, that there's a lot of interesting philosophy and science and technology for us to do in this field; if not literally tomorrow, then anyhow quite soon. And even if *we* don't get to do it, we can create the organizational framework in which our graduate students will. This is, insofar as we care about the advancement of the scientific enterprise, not just our opportunity but also our obligation.

Research Goals in Information, Computation and Cognition

Barbara H. Partee

The Turing Test as an Interdisciplinary Goal of Basic Theoretical Research in Information, Computation and Cognition

For a machine to pass the Turing test, suggested by Alan Turing, it has to succeed in carrying on a conversation with a human partner in a way indistinguishable from another human. Getting machines to pass the Turing test, whether ultimately possible or not, can be taken at least metaphorically as one long-range goal of AI: but it can be equally taken to be, indirectly, a long-range goal shared among all the fields of AI, computer science, linguistics, logic, philosophy, and psychology. Many researchers give at least serious lip service to the idea that if one really understands some domain of human cognitive ability, one would in principle be able to program a machine to exemplify it, at least if one's colleagues could do the same for all the other abilities the ability in question interacts with. In syntax, for instance, virtually all work in modern times accepts such a goal; a central aim of generative grammar is to make the unconscious knowledge of the native speaker explicit, and "explicit" is generally identified with "in principle programmable." Actual programming is of course not demanded; being able to formalize a grammar with finitely many primitives is generally regarded as relevantly equivalent. In fact in syntax, and in linguistics generally, the goals are set even higher; no one would be content with even a machine whose command of English was indistinguishable from ours, or even machines with perfect Dutch, Bengali, Korean and Kwakiutl. What the linguist demands is a "child" machine that starts out with no particular language and succeeds in learning whatever language the adult humans (or machines) around it are speaking, with no more explicit instruction than a human child gets. We could call this the "baby Turing test" for linguistic ability. Distant as such a goal may seem, its recognition as a goal, at least metaphorically, could fairly be said to have marked the beginning of the tremendously explosive progress we have seen in generative grammar in the last several decades. Linguists are primarily interested in human abilities, not machine capabilities; but insofar as they demand of their theories this

kind of explicitness, a successful theory should be able to help AI researchers get their machines to pass the Turing test (and the harder baby Turing test).

Of course the behavioral nature of the Turing test has its dangers, as philosophers have pointed out. But in serious research, I believe we really are entering a stage where good AI work and work in theoretical computer science are giving linguists, cognitive psychologists, philosophers and logicians new and interesting hypotheses about human abilities as well as those sciences giving AI researchers and theoretical computer scientists useful ideas for their enterprises, and progress from both ends toward the middle, so to speak, is bringing convergence of opinion that there are urgent and difficult problems which require concerted attack by interdisciplinary teams of researchers. It is as if we had been working from both ends trying to dig a tunnel through a mountain, not even sure we were in the same mountain, and now we can hear each other working, have confidence that we are converging, have all discovered that the rock in the middle of the mountain is the hardest, but together are ready and eager to assault it and have already opened up chinks through which we can lend each other our best tools, which then come back to us with unanticipated improvements.

Such research is opening up exciting prospects for major advances in foundational aspects of pure science as well as holding promise of long-term benefits for applications of great importance. I believe it is widely understood that the computer and communications technology of the next decade depends crucially on the scientific advances of this decade.

Some Fundamental Research Questions

Now I'll try to give my own view of some of the fundamental research questions that are emerging in this convergence in cognitive and information science. I'll state them in rather broad terms; these are meant as examples and not as an exhaustive or definitive list.

The Relation Between Logics and Reasoning by Finite Agents

Logic as studied by logicians has always been a kind of science of the ideal; it is concerned with what conclusions *should* be drawn from given premises, what conclusions really follow. The study of how humans actually reason is a separate matter, of interest both as a basic question about how the mind works and as a potential source of insight for designing intelligent machines. The computer age has created an important bridge between these concerns and has had exciting effects, since it introduces in a third question: what's the best reasoning that a finite machine can do? In the course of debates on what one should optimize and research on how to optimize various prop-

erties of such a system, old logics have gained new life and new logics have been spawned. New respect has been gained for human reasoning, where one sees now not just susceptibility to error but tremendous adaptiveness in being able to reach reasonable if not infallible conclusions, in being able to revise intelligently beliefs when errors are detected, and in being able to transfer structure and patterns of reasoning from one domain to another, in what computer scientists call "transportability." There are the beginnings of a creative outpouring in this domain, together with a clear need to bring researchers from different fields together to benefit from their different perspectives on this family of problems, as well as from their different intellectual tools and achievements.

Bringing the Language User into Semantics

Formal semantics, which was developed by logicians and which has been a vital part of the remarkable recent progress in semantics of natural languages, idealizes away from the language user and treats meaning as a relation between language and the world (or other models); the meaning of an expression is identified with its contribution to the truth-conditions of sentences it occurs in. Formal pragmatics brings a language user into the picture and allows one to consider context-dependent expressions like I, here, and now, but has generally treated the language-user as an ideally rational agent with infinite logical capacity. Application of these theories to human and computer languages has been extremely fruitful, but some of the hardest problems, such as the problem of propositional attitudes, are leading many researchers to the conclusion that the finiteness and consequent limitations of human language users has to be taken into account in an essential way in understanding how human language works. Recent progress in the semantics of computer languages (where the limitation of finiteness has always been a central concern, and where attempts are being made to develop languages with expressive power closer to that of human languages) is clearly resulting in both the need and the possibility of richly productive collaborative efforts. The contributions of philosophers and logicians have been and will continue to be as crucial to the success of this enterprise as those of linguists and theoretical computer scientists. And if progress can be made in developing realistic processing models for semantics and pragmatics, we can expect to build stronger links with research in psychology and in AI as well.

Giving Up the "Closed World" Assumption in Knowledge Representations

Question-answering systems that are connected to data bases are typically designed so that when a yes-no question is asked, the machine tries to ver-

ify the corresponding assertion: if it succeeds, it answers "Yes," otherwise it answers "No." But sometimes it should in principle answer "I don't know." It turns out to be quite a difficult problem to design a system that knows when to say "I don't know." The "closed world" assumption amounts to the dogmatic position "If I don't know it, it isn't true." Related problems arise for the semantics of negation in partial models, for the semantics of the programming language PROLOG which equates negation with failure, and in the analysis of propositional attitudes, where the distinctions among disbelief, agnosticism, and total absence of attitude can only be made in the context of powerful systems for reasoning about one's own beliefs. Awareness of one's own possible ignorance makes life infinitely more troublesome; but I believe it is one of the most significant among distinctively human attributes—it is, after all, what makes science possible, among other things. Computers that had some "awareness" of their limitations could be not only more trustworthy, but certainly more user-friendly. This is an area with rich promise for both theoretical and practical payoff.

Developing a Dynamic and Integrative View of Language Understanding

If we have to abandon the view of meaning as (static) truth-conditions and bring the language user more centrally into the picture, we also have to see the process of language understanding as both affected by and affecting the context of language use and as crucially involving its integration with an effect on knowledge, belief, and other attitudinal states of conversational participants. Conversely, the perennial issues of analyzing the nature of such mental states takes on a new light if one takes as central the question of how beliefs, etc. change in the face of incoming information.

I hope it's fairly clear that all of the issues I've mentioned so far actually tie closely together. One common thread of concern is the nature of the information that an information processing system, machine or human, deals with. It is becoming clear that an adequate account of this notion requires simultaneous consideration of a language, a language user (human or machine), and the world, facts, situations, or whatever the language purports to be about.

Replacement of idealizations by more realistic assumptions brings previously partial and incommensurate analyses of phenomena in the emergent problem domain into direct contact. It's hard to give up idealizations without making resulting problems hopelessly messy, but if the idealizations have been in a sense complementary and the results made with them are robust, researchers can fill in many of each other's gaps and jointly tackle the hard problems that no approach individually has been able to solve.

A Shift in Emphasis in the Study of Representation

In the study of *representation*, including knowledge representation, linguistic levels of representation, and visual and other forms of representation, I believe we are seeing a shift of focus from concern primarily with the existence (or "psychological reality") of systems of representation and their formal properties and from the nature of purely symbolic processing systems to a concern with how representations come to have content and the question of the semantic relation of representations to whatever they are representations of, for both humans and machines. And then of course we confront the difficult intensional questions that arise when representations seem to function as if they were of something but in fact are not, as often arises in cases of false beliefs. As Bob Ritchie has noted, this issue is becoming acute in computer science in the networking of machines with very different internal workings, where it is no longer possible to take the semantics of their languages as being defined simply as their operational effect, and what the expressions of their interface language mean starts to take on the same difficulties as corresponding questions for hypotheses about, e.g., the "language of thought" that figures in representationalist approaches to natural language semantics.

Replacing Set Theory by Property Theory, Developing More Intensional Notions of Function, and Related Issues

This is a family of issues on which I'm not an expert, since the main developments are taking place within theoretical computer science and mathematical logic, but I want to include them because I'm convinced they are needed for progress in semantics of propositional attitudes and for the semantic account of any language used by finite agents with limited computational resources. Procedural semantics is an attempt to move in this direction and needs better formal foundations. Intuitionist logic is gaining new relevance in this context, partly because it embodies a much more constructive notion of proof than classical logic. Self-application of functions, important in both natural and computer languages, is possible in most versions of property theory but not in most versions of set theory. It is also possible in Dana Scott's domain theory, developed out of Church's lambda calculus as a means of providing a formally rigorous semantics for computer languages. These developments seem to me crucial for providing foundations for semantics that have the power and soundness of systems like Montague's intensional logic (which turns out not to be intensional enough) which are compatible with the basic finiteness constraint of human and machine language users using richly expressive languages. Much of the relevant exper-

tise and active research has arisen within mathematical logic and theoretical computer science: collaborative efforts between these fields and studies of human language are just getting under way, and logicians are just beginning to explore ways of incorporating insights from empirical linguistic research in the design of foundational systems of logic and semantics, with early results of great promise for all the contributing fields. The problems are very difficult and the communication gap between the directly relevant fields is still a serious impediment to rapid progress. There is a marked scarcity of people with relevant knowledge and interests, and a pressing need exists to attract both senior and junior researchers into this area and provide environments in which, say, logicians working in theoretical computer science and linguists can gain enough mastery of each other's fields to work together on these problems.

The Time Is Ripe

Interdisciplinary research always requires extra effort and extra resources of time and money, so a major thrust of the sort we have been suggesting should certainly not be undertaken lightly. Premature attempts can actually set a potentially valuable enterprise back and sour the research community and the sources of support on its viability, as perhaps happened with the push for machine translation in the 1950s. Individual disciplines need idealizations to study specific facets of any big hard problem area, and premature attempts at integration or application of theories in young disciplines are likely not to be fruitful. (Alongside the truism "Garbage in, garbage out," there should be one about the exponential explosion of garbage when you try to multiply.)

It would be misleading to suggest that all the separate enterprises reflected in the converging subparts of the six or more disciplines represented in our meeting are *settled* and we're just ready to *integrate*. But I feel that there are strong grounds for optimism about the success of such a venture now because one senses a readiness to make further progress by modifying the simplifying assumptions in our respective subfields in ways that seem *convergent*. And at this point the pressures for reaching out across disciplinary boundaries are as strong from within the separate fields as they are from our perceptions of the long-run potential such joint work could have for the development of what might turn into a new discipline if it is successful. The Sloan Foundation and the System Development Foundation gave important and far-sighted encouragement to interdisciplinary work in two overlapping subparts of the domain we have been trying to define; I think we have now crossed some threshold where significant parts of the research community, as well as the leaders of those industries that are responsible for new technological developments in this area, see this kind of interdisci-

plinary attack on central hard problems as a necessity for future progress rather than as just an intriguing prospect that they will take the effort to explore if they can find the time and resources. But consensus about priority and urgency, while crucial—and I believe our meeting made it clear that the consensus is deep and strong—is not enough by itself to turn the potential into reality. That will take substantial resources, with stability to make long-range large-scale research projects possible, and with the involvement of an established national funding agency such as the NSF to facilitate the involvement of researchers from institutions large and small all over the country, the establishment of additional computer net facilities to link such researchers together, and the peer review process which is so essential to the progress of science. Without such resources, progress in this area will be much slower, fragmented, and haphazard; with them, I believe we will see an explosion of exciting groundbreaking research with immense significance both for this newly developing field together with all its contributing disciplines and for the next generation of computer and information technology.

A Presentation to the National Science Foundation Workshop on Information and Representation

Mitch Marcus

The speakers before me have described some recent intellectual trends and scientific developments which point to one result: There is a newly emerging approach to the scientific study of the interrelation of mind, language, and the world. This study is characterized in part by the crucial drawing together of methodologies and tools from a wide range of existing disciplines, as well as a startled recognition that problems and solutions in one discipline are often closely linked to those in others.

I'd like to underscore the fundamentally interdisciplinary nature of this undertaking by focussing on one set of interactions: those in which concepts and analytic tools from computer science play a crucial role. That computer science is centrally involved shouldn't be surprising; the mind is often viewed as a very special kind of processor of formal symbolic representations. Central to computer science is the study of the formal structure of symbol manipulating processes, the formal properties of mathematically characterized classes of languages, and the semantic characterization of programs, i.e., what programs "mean." All these are closely parallel to related studies of natural language.

Let me begin by briefly running through a set of some parallels. I should stress that these examples are all over-simplified for the sake of brevity. (See Fig. 1.)

Let's begin with syntax. To a first approximation, the grammars of both natural languages like English and French and the artificial languages of programs can be characterized by the same formal tools. For example, take the phrase shown in Figure 1a, a phrase of some programming language. Its grammar can be represented by what are called *phrase structure rules*; two relevant phrase structure rules are given in Figure 1b. The first says that an Exp (i.e., an expression) can be made up of the token "while," followed by a bool (i.e., some boolean expression), followed by "do," followed by another Exp. A bool itself is made up of a term, followed by a boolean predicate, followed by another term. Sentences in natural languages can also be represented by such rules. Take for example the sentence "The dog bit the man." (Fig. 1c). Its structure can be described by the rules shown in Figure 1d, which says, An S (sentence) is made up of an NP (noun

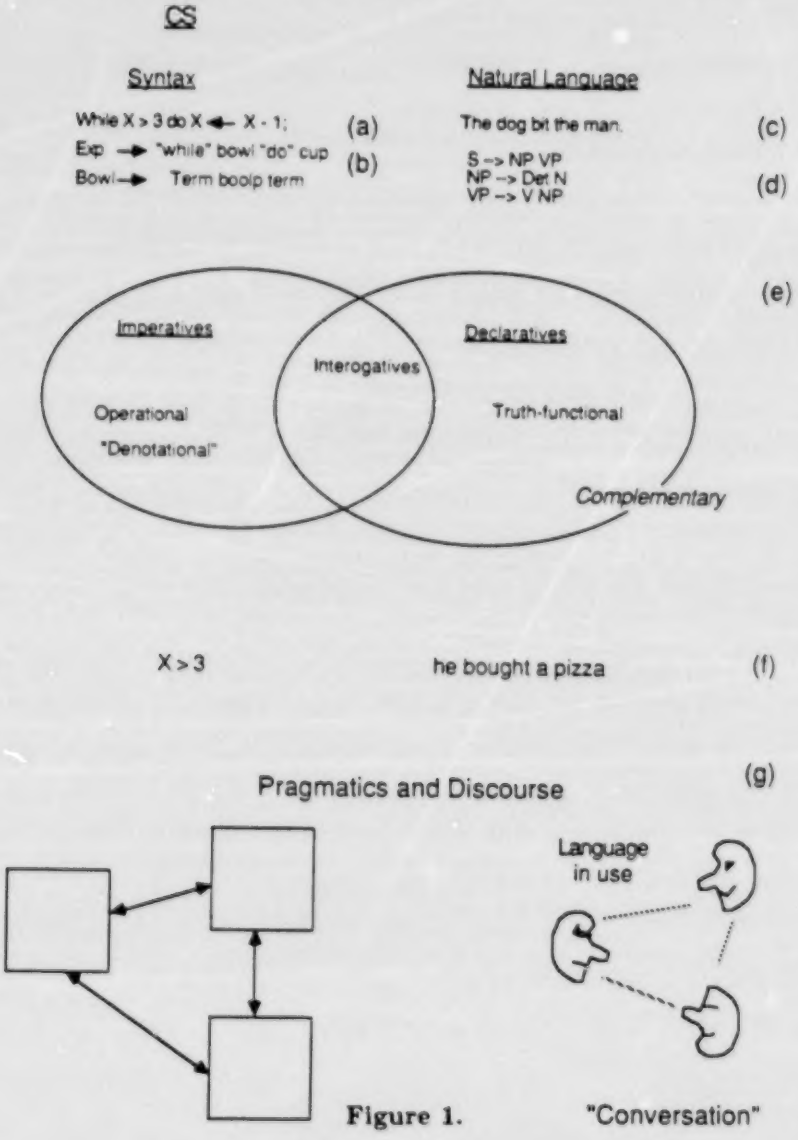


Figure 1.

phrase) followed by a VP. An NP can be made up of a Det (like "the"), followed by a noun (like "man" or "dog"), and so on. As we will see in a moment, however, the syntax of natural languages requires extensions to this simple notion, but the point here is that the same formal notion serves as a starting point in both cases.

Computer science, like linguistics and philosophy, has wrestled with the problems of semantics—precisely formulating the meaning of program texts on the one hand, and natural languages utterances on the other. What is interesting is that past studies of the formal semantics of programming and natural languages have been largely complementary. (See Fig. 1e.) Computer programs are by and large *imperative*: i.e. orders (for a machine, in this case) to *do* something. To formally specify the meanings of these imperatives, computer science has attempted to formulate just what abstract function a program computes (which computer scientists call "denotational semantics") or exactly what it is that the execution of a program does ("operational semantics").

While NL uses imperatives, like "Shut the door" or "Don't come home too late," work on the semantics of NL has focussed centrally on *declaratives*, sentences that say what is the case, e.g., "Two plus two are four" or "John just arrived." Classical semantic models developed for declarative sentences have been largely *truth-functional*: i.e., they attempt to characterize the meaning of assertions by relating the assertions to those circumstances in which a given assertion is true.

The point here, of course, is that as we increase our understanding of both natural and programming languages and augment the kinds of computer languages we have, we will need both kinds of semantic analyses coming together; we will need some kind of synthesis. Furthermore, interrogatives, realized as questions in natural language and as data base queries in computer science, have semantics that are slightly different from imperatives and declaratives. What we really need is some theory that will unify all these statement types into one simple framework; interdisciplinary work of this kind is already going on, but these frameworks are too technical for a simple exposition.

Another issue in semantics of crucial importance is what referring expressions mean (see Fig. 1f). So, in a computer language, you may have fragments like "$X>3$," and the question of course, is what do you mean by "X," what is X's value? Or, "*he* bought a pizza," and the question here is "what do you mean by he?" As the other speakers here this morning have indicated, this is a central problem; we will investigate some computational implications of this below.

Other areas include the study of pragmatics and discourse, or perhaps somewhat more generally, language in use. The crucial point here, emphasized by John Perry as well, is that conversation is in fact a flow of information between entities. In computer science, we talk about this as distributed processing (Fig. 1g), an area that is just beginning to develop and about which we really don't know very much. Some recent work in com-

puter science, which I will review below, suggests that some of the things we have learned from looking at conversation, and at the prerequisites for conversation, are exactly relevant to distributed processing.

One important pay-off of being able to really understand the way natural language works is that we may be able to come up with formal languages that approach the brevity and conciseness of natural language. Figure 2 shows a query automatically translated into a data base access language. This particular example is taken from the LADDER system done at SRI International, but many other examples could have been chosen, with the same general result: The translation of an eight word query is 26 lines long, including declarations, and includes three nested loop constructs. If we could formally understand how it is that *we* understand the simple eight word query above, we just might be able to come up with formal query languages and perhaps ultimately programming languages that exhibit the same conciseness.

What follows will focus on three exemplars of the interaction of computer science with the study of natural languages. In particular, I will discuss the following:

1. Formal theories of grammar—the interchange between abstract mathematical models of formal grammar and models of natural languages.
2. The notion of discourse structures—the relationship between discourse structure of the kind that John Perry also discusses and the block structure constructs of ALGOL-like programming languages. As it turns out, there is a deep similarity.
3. The notion of common knowledge or mutual belief—one of the prerequisites for two individuals to talk to each other felicitously. When I talk to you, I can use a wide range of referring expressions and you know what I mean as long as I depend upon descriptions that we each know the other knows, and we each know that the other knows that we know. It is this notion that has led to some surprising results in the area of distributed computation.

As a final point, I will briefly mention the notion of *semantically rational* programming languages. Some recent work on this leads naturally to the implementation of entirely portable debugging environments of great power. The interpreter of a semantically rational language can be viewed as operating on semantically sensible expressions that *refer* to entities in the language itself, rather than manipulating arbitrary structures that depend on the particular computer or particular language being used to implement the target language. This allows the language to include constructs that allow access to the inner (virtual) structure of the interpreter itself which

A LADDER query and query translation

English Query:
To what country does the fastest sub belong?

DATALANGUAGE Query:

```
BEGIN
DECLARE Y1 STRING (.100) ,0 = '}'
DECLARE Y2 STRING (.100) ,0 = '}' Y2 = '00.0'
DECLARE Y3 INTEGER Y3 = 0
DECLARE Y5 STRING (.100) ,0 = '}' Y5 = 0
DECLARE Y4 STRING (.100) ,0 = '}' Y4 = 0
FOR R1 IN SHIPCLASCHAR WITH (R1. TYPE2 EQ 'S')AND(R1.TYPE1 EQ 'S')
   FOR R2 IN SHIPCLASDIR WITH (R2.SHIPCLAS EQ R1.SHIPCLAS)
      FOR R3 IN SHIP WITH (R3.UNICVCN EQ R2.UICVCN)
          BEGIN
          Y1 = R3.MCSF
          IF Y1 LE '99.9' AND Y2 LT Y1 THEN
              BEGIN
              Y2 = Y1
              Y5 = R3.NAT
              Y4 = R3.NAM
              Y3 = 1
              END
         END
IF Y3 EQ 1 THEN
    BEGIN
    MSTDPORT.STRING1 = Y4
    MSTDPORT.STRING2 = Y5
    END
END
```

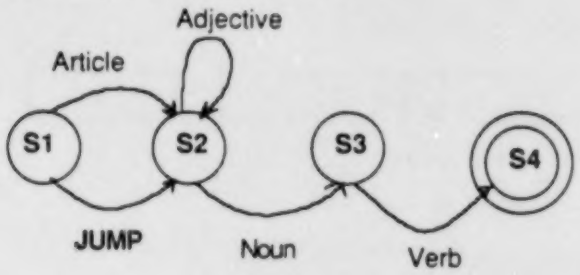
Figure 2.

is operating on this invariant referential structure. These constructs then allow the quick implementation of debuggers that move transparently to implementations on any other machine at all, because the debugger itself is based upon the virtual semantic structure alone.

Turning to the first case study, we will now examine some work which studies the formal properties of the syntax of natural languages. (This discussion is necessarily more technical than the case studies that follow.)

A good starting point historically is the late '50s when notions from computer science and linguistics interacted for the first time. There was a real explosion of work in formal language theory, including much done by

BEST COPY AVAILABLE



(a)

(b)

Transition networks accept regular grammars.

(art + Ø) adj* noun verb

(c)

Static vs. Dynamic

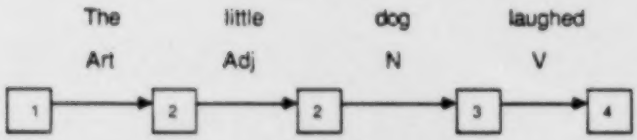


Figure 3.

the linguist Noam Chomsky; I would like to give you a taste of this work as it lays the basis of what follows about 20 years later.

Chomsky's work posits a computational hierarchy of formal languages. At the bottom of this hierarchy are what are called *regular grammars*; these can also be modeled by a construct called a *transition network*. (See Fig. 3a–b.) A transition network consists of a set of states and a set of transitions between those states. The transition network shown in Figure 3a and related regular grammar shown in Figure 3b describe the simple artificial language that consists of an optional article, like "the," followed by any number of adjectives, possibly none, followed by one noun, and then one verb.

There are two related but different ways of viewing such a model. One can view such a model *statically*, or one can view it *dynamically*. Viewed statically, the transition network given here is a purely mathematical entity that defines a language which consists of a particular infinite set of sentences, namely those sentences which begin with an article, then any number of adjectives, then... Viewed dynamically, this network serves as a *program* (for a particular kind of simple abstract computer) that when executed will recognize word strings that are sentences within the language the network describes. This machine has a register which simply keeps track of which

of its states the network is in at any particular instant. Given an input like "the little dog laughed" (see Fig. 3c), the machine starts in S1 which is the starting state, "the" takes the machine to S2, "little" takes it back to S2, "dog" takes it to S3, and "laughed" takes it to S4. S4 is a special final state, as indicated graphically. Since the input is exhausted while the machine is in a special final state, this string is a sentence in the language the transition network describes.

So one can view this kind of model as statically defining an infinite set of languages or as providing a dynamic *incremental* process for recognizing sentences. As it turns out, one can augment such a dynamic model just a little to provide a *parser*, a mechanism that outputs an analysis of the grammatical structure of input sentences.

We now take one step up the Chomsky hierarchy towards the kind of computational richness required for natural languages: One can allow subnetworks of such a transition network to be extracted and named, allowing the transitions to consist of "subroutine calls" to these subnetworks, as shown in Figure 4. Such a set of networks, taken together, makes up a *recursive transition network* (RTN). The toplevel network in Figure 4a, encodes that a sentence consists of a successful transition through the noun phrase network (Fig. 4b), followed by a successful transition through the verb phrase network (Fig. 4d).

To view this model dynamically, the abstract machine which executes these RTNs must use a *stack* as well as a state register. Consider the example network of Figure 4. The NP network calls the PP network, *which itself calls the NP network recursively*. So the machine must keep track of where it was in the higher networks to keep track of what its doing, i.e., it must keep a stack of "subroutine calls." So, given "the little dog laughed" and this grammar, the machine will start out in state S1, then call the noun phrase network. When it returns, it will be in state S2, so this is saved on the stack (see Fig. 4e). Now in the NP network, "the" takes the machine into NP2, and the adjective "little" takes up back into S2. At this point the stack has the return state S2 in it, and the register stores the current state NP2, as shown in Figure 4e.

Now it turns out that the general structure of natural language syntax cannot be formalized within this RTN framework; a richer formalism is required. Two simple examples (see Fig. 5): a sentence can be made up of a *singular* noun phrase, followed by a *singular* verb phrase, or of a *plural* NP followed by a *plural* VP, as shown in Figure 5a. If you represent this within a simple RTN, one needs two NP subnets and two VP subnets; there is no simple way to express the fact that a sentence consists of an NP followed by a "matching" VP. Another problem: It turns out that natural language

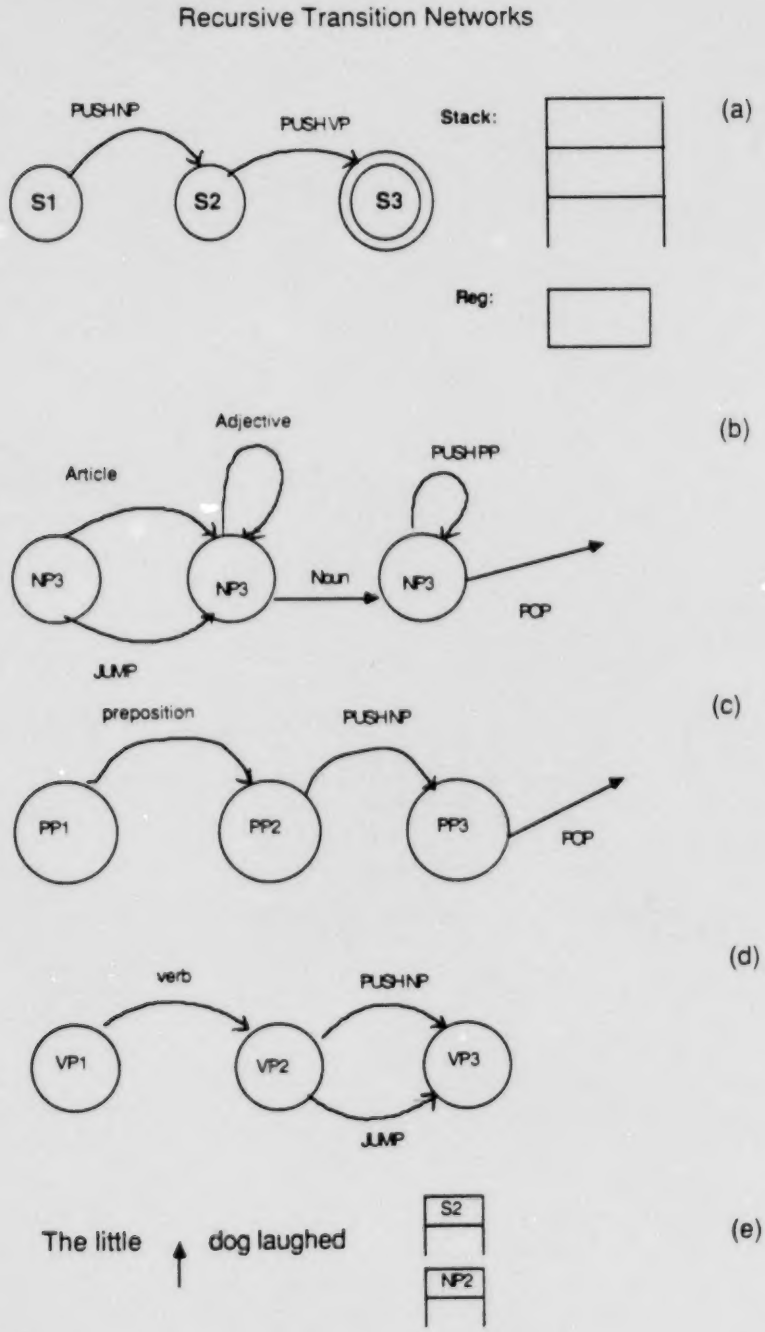

Recursive Transition Networks
PUSHNP
PUSHVP
S1
S2
S3
Stack:
(a)
Reg:
Adjective
Article
PUSHPP
NP3
NP3
Noun
NP3
POP
JUMP
(b)
preposition
PUSHNP
PP1
PP2
PP3
POP
(c)
verb
PUSHNP
VP1
VP2
VP3
JUMP
(d)
The little
dog laughed
S2
NP2
(e)

Figure 4.

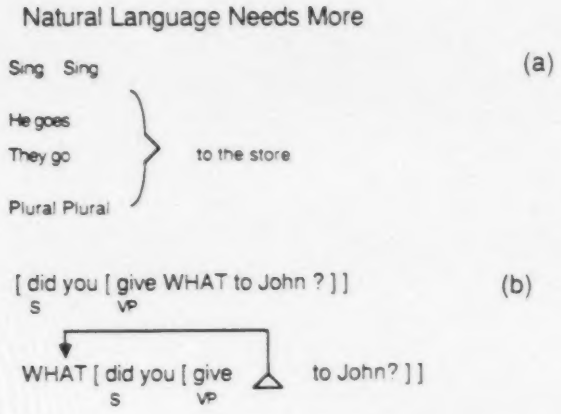


Figure 5.

sentences can have "holes" in them. So the sentence

WHAT did you give to John?

can be viewed, in some sense, as deriving from something like

Did you give WHAT to John?

where "WHAT" gets pulled to the front, if you will, leaving something like "what did you give to John?" where there is a hole that represents the thing that was pulled out in some sense or other. (See Fig. 5b.) It also turns out that these holes can be arbitrarily far away from the place in the sentence where "WHAT" appears on the surface. The simple RTN framework (the equivalent of what Chomsky called *context free grammar*) is also incapable of capturing this phenomenon without further extension.

Having developed this computational framework, Chomsky now made two key intellectual moves. First, he embraced fairly solidly the static view of language, thereby removing the study of linguistics by and large from processing considerations for about 20 years, although there were a few die-hards who kept thinking about things in dynamic terms. Second (but relatedly), although Chomsky developed another two steps on this computational hierarchy, he opted to remain within an earlier, rather different formal model based on string algebras on which to base his revolutionary formalization of the grammar of natural languages. One crucial property

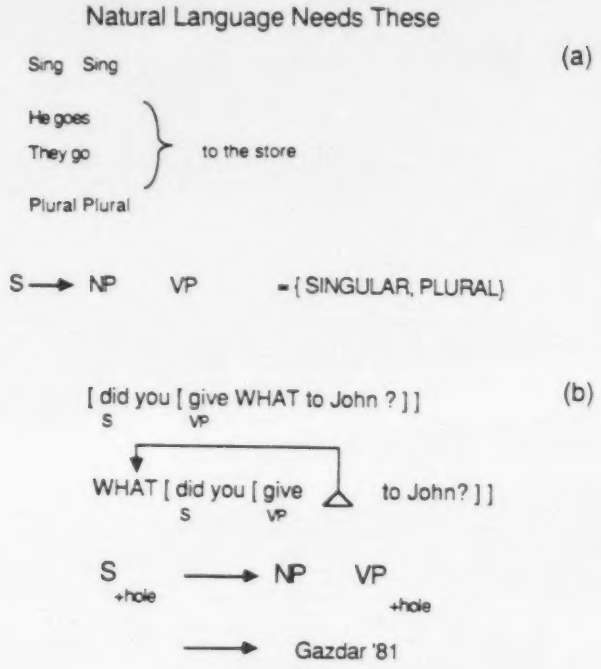


Figure 6.

of this string algebra approach is that it suggests no simple incremental parsing interpretation, thus strengthening the gap between linguistic theory and computational practice. With a few exceptions, this is more or less how things remained for twenty years.

In the late '70s, struck by some mathematical results in formal language done by Joshi and Levy, extending earlier results due to Peters and Ritchie (it's worthy of note that three of these four are present at this workshop), Gerald Gazdar realized that the set of phenomena presented above could in fact be elegantly expressed within formalisms which could be algorithmically expanded back into the RTN formalism; i.e., that it looked promising that natural languages were within the second step of the Chomsky hierarchy.

Let me try to give you a sense of the formal devices Gazdar used: Gazdar used a particular feature notation with feature variables. This allows the statement that a sentence is made up of a noun phrase with a number feature variable alpha, and a verb phrase with the same number feature alpha, where alpha is either singular or plural, thereby capturing the generalization about subject-verb agreement (see Fig. 6a). He introduced a formalism that captured the fact that sentences with initial WH-phrases have "holes" where

the WH-phrase would be in a simple declarative (i.e. "WHAT did you give ⟨hole⟩ to John?") by allowing sentences to have a "hole feature" associated with them, which is transferred to some subconstituent, in this case to the verb phrase (see Fig. 6b). As a result of this work, other phenomena of natural language have come to light that seem to resist formalization within notations that can be mapped back to RTNs; Gazdar and his collaborators have now moved to somewhat more powerful computational models. But the key point here is that some deep theorems about the nature of a formal mathematical system gave rise to important innovations in characterizing the structure of natural languages that led to an important increase in our knowledge about the syntactic structure of natural language.

Another example is the dual of Gazdar's work: In an attempt to study in isolation the formal mathematical structure of the class of syntactic operations which extends beyond the power of RTNs, Joshi has developed a new formal language model which he calls tree adjoining grammars (TAGs). The TAG model can be used to generates sentences with "holes" arbitrarily far from the associated phrase (as in Fig. 7a), "Who did John persuade Bill to invite ⟨hole⟩?" by generating all holes within limited contexts (Fig. 7b), and then "tearing apart" some particular node (in this case an S node), and adjoining an independent subtree which separates the hole from the WH-phrase by an arbitrarily large distance (Fig. 7c). This TAG model lies above the RTN step of the Chomsky hierarchy, but is provably much weaker than the step on the Chomsky hierarchy above it; in some sense, it is only slightly more powerful than the RTN model.

At about the same time that Gazdar and Joshi were proposing these new models, other researchers began to reject Chomsky's view that linguistic investigation be based upon a purely static view of language (Fig. 8). The notion of "psychological reality," that a linguistic model must be amenable to some range of process interpretations, has recently given rise to a wide range of new linguistic models. My own work, for example, has been based on the working hypothesis that a grammar should be viewed as a program written in a very special kind of programming language, where the properties of natural language must derive from the execution of that programming language by some particular universal mental interpreter. The structure of this interpreter itself, under this view, is the locus of properties which are common to all human languages, which collectively make up what Chomsky calls universal grammar.

The work of Bresnan and Kaplan and their colleagues has attempted to have the best of both worlds, with rich results so far. They have proposed a very powerful formalism which, like the RTN formalism, can be viewed both dynamically and statically. Linguists can work within this formalism

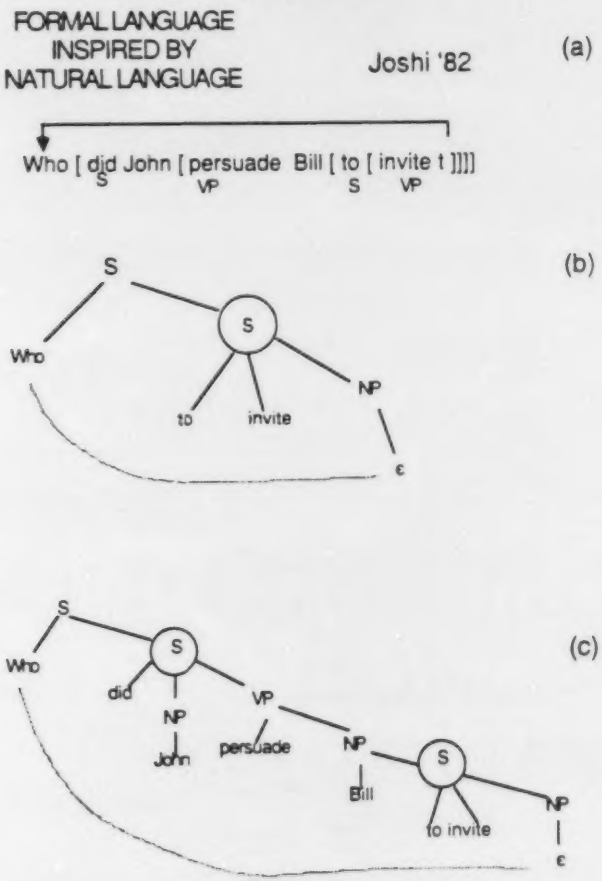


Figure 7.

to create static models of the grammars of particular languages, fine-tuning the formalism to reflect universal properties of human language, in much the same way that linguistics has been done in the last twenty years. But at the same time psychologists can view this formalism dynamically, looking for those interpreters that account for the human abilities to parse language, to generate sentences, and to learn to do this within childhood. In a more practical vein, computer scientists can provide interpreters that run efficiently on existing computers, allowing the linguist's grammars to be utilized within natural language interfaces to existing computer systems.

Let me turn now to my second case study: I want to discuss deep commonalities in the interpretation of referring expressions within both natural and programming languages.

BEST COPY AVAILABLE

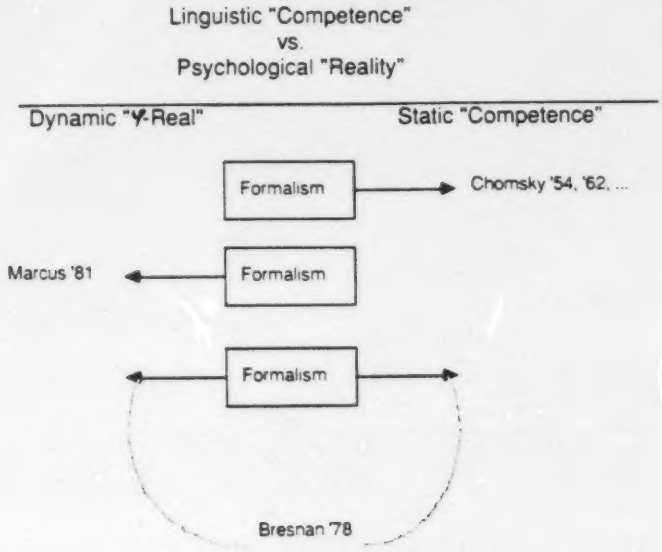


Figure 8.

Block Structured Programming Languages

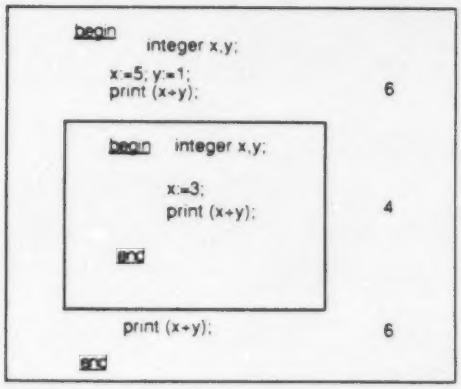


Figure 9.

Figure 9 shows a fragment of ALGOL code, exhibiting two nested blocks of variable declarations. In the outer block, X is bound to 5 and Y to 1. The side effect of executing "print$(x + y)$" in this block is that "6" gets printed. In the inner block, we rebind X to 3, and execute "print$(x + y)$" again. Within this inner context the result is that "4" gets printed. But after the inner block is exited, the execution of the same expression results, once again, in the printing of the numeral "6." This is all quite commonplace to computer scientists.

More generally, variables occur within nested blocks, and variables which are rebound within inner blocks retain their new bindings within the scope of the block. The surprise is that the change of reference that John Perry describes in human discourse exhibits *exactly the same formal behavior*! Figure 10 shows schematically a fragment of a real monologue (changed just slightly, actually). Reichman, building on earlier work by Grosz, notes that at the beginning and end of the fragment "he" refers to the speaker's cousin, *both before and after a long digression*, within which "he" refers to the speaker's father. So formally, the digression serves as an inner block, within which the binding of the pronoun "he" changes. But upon returning to the outer block, the reference of "he" reverts to its assignment within that block. Just as in ALGOL!

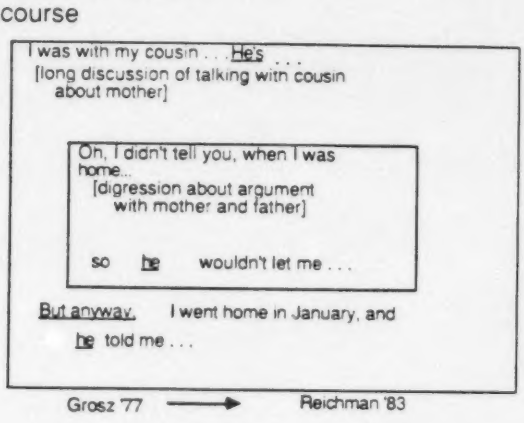

Discourse
I was with my cousin . . . He's . . .
[long discussion of talking with cousin about mother]
Oh, I didn't tell you, when I was home...
[digression about argument with mother and father]
so he wouldn't let me . . .
But anyway, I went home in January, and
he told me . . .
Grosz '77
Reichman '83

Figure 10.

What's really going on here, Grosz and others have argued, is that the the topic structure of a discourse, as well as the variable binding structure of an ALGOL-like programming language, is maintained on a stack. (See Fig. 11.) Executing an ALGOL program, the machine maintains a stack of variable bindings. When the inner block is entered, the machine pushes a new binding, with $X = 3$, onto the stack, hiding the old binding, where $X = 5$. When the machine pops back to the outer context, lo and behold, $X = 5$ again. Similarly, according to Grosz and others, when we talk, just as when we perform many other activities, we have a stack of attentional contexts. In this monologue, for example, the attentional context is initially focussed on the speaker's cousin, then there is a "push" to a subcontext focussing on an argument with her parents, and finally attention shifts back to the earlier context. The same formal device, contexts pushed and popped from stacks, accounts for the reference of referring expressions in both the

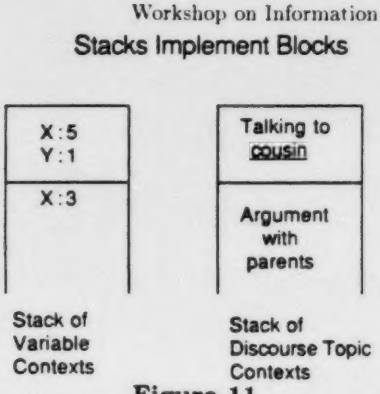


Figure 11.

human and the artifact. The parallels here can be continued to much deeper levels, but this much illustrates the central point.

One more illustration of this point, just to provide evidence of its generality: Consider the dialogue fragment shown in Figure 12, a conversation between an apprentice learning how to assemble a simple compressor and an expert. [Quick digression: A wheel puller is a tool with two side arms that grasp the edges of a wheel, with a central screw which pushes against the axle, pulling the wheel itself forward off the axle.]

A: How do I remove the flywheel?
E: First loosen *the two small allen head setscrews* holding it on to the shaft. Then pull it off.
A: The *two setscrews* are loose but I'm having trouble getting the wheel off.
E: Use the wheel puller. Do you know how?
E: Loosen *the screw* in the center and place the jaws around the hub of the wheel, then tighten the screw.

Figure 12.

Now the crucial fact here is that the phrase "the screw" *does not* refer back to one of these two set screws, even though they've just been discussed, but rather to the central screw of the wheel puller. And we followed without any confusion. How? Because we automatically track the topic structure, and note a "push" in the topic from the flywheel to the wheel puller. Within this subcontext, the binding, so to speak, of "the screw" has shifted. The formal structure all follows from the notion that the same computational device, a stack, underlies a wide range of focussed behavior.

Let's turn now to the third case study, the role of common knowledge, of mutual belief, in people and processors. Consider the following example, taken from work by Clark and Marshall:

> On Wednesday morning Ann reads the early edition of the newspaper, which says that *A Day At The Races* is playing that evening. Later she sees Bob and asks, "Have you ever seen the movie showing at the Roxy tonight?"

The question on which I'd like to focus is simply this: In discourse between individuals, what does the hearer have to know and what does the speaker have to know to make natural and appropriate the use of the expression "the movie showing at the Roxy tonight?"

The answer is both subtle and surprising, as the following variant makes clear; it turns out that it is far from sufficient for both Ann and Bob to know what's playing at the Roxy tonight:

> On Wednesday morning, Ann and Bob read the early edition of the newspaper and they discuss the fact that it says that *A Day At The Races* is showing that night at the Roxy. When the late edition arrives, Bob reads the movie section, notes that the film has been corrected to *Monkey Business* and circles it with his red pen. Later, Ann picks up the late edition, notes the correction and recognizes Bob's circle around it. She also realizes that Bob has no way of knowing that she has seen the late edition. Later she sees Bob and asks: "Have you ever seen the movie showing at the Roxy tonight?"

If you think about it for a moment, it becomes clear that Bob must think that the phrase "the movie showing at the Roxy tonight" used by Ann refers to *A Day at the Races*, even though they both know that the movie showing is in fact *Monkey Business*. What do they each have to know to get it right? Clearly, more than they do here. What do they know? Well, Ann knows P, where P here is the fact that *Monkey Business* is playing, and Bob knows P, that *Monkey Business* is playing. Furthermore, Ann knows that Bob knows P. But crucially, if you think about it, Bob doesn't know that Ann knows P. Further similar examples show that even if Bob knows Ann knows P as well, he would have to know that *she* knows that he knows that she knows P, and so on.

It turns out that somewhere down this hierarchy of Ann knowing that Bob knows that Ann knows that Bob knows... seems to lie the answer. In fact, it has been convincingly argued by Lewis and Schiffer that the answer lies exactly in the infinite extension of this hierarchy, and that this is what constitutes mutual belief or common knowledge, that I know that you know that I know that you know and so on, and so forth for ever. What Clark and Marshall show is that this mutual belief is a prerequisite of the simplest kinds of understanding referring expressions.

But how could this infinite hierarchy be represented in finite brains or finite computers? Cohen, who independently discovered Clark and Marshall's

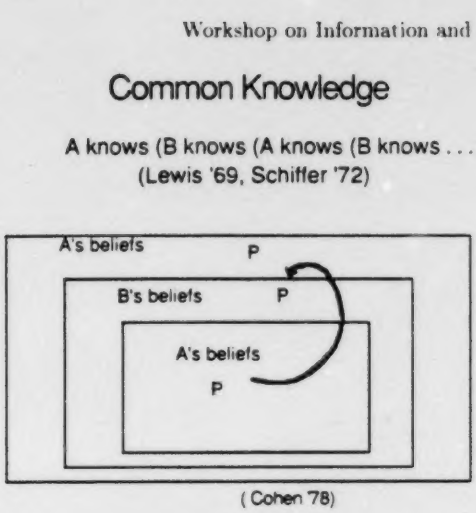


Figure 13.

result, also suggested a simple data structure using the notion of *pointers* to solve this problem. (See Fig. 13.) Consider a data structure which represents the set of Ann's beliefs, somehow nested in Ann's head, if you will, and in that set is P. But Ann knows some of Bob's beliefs, so within this data structure representing Ann's beliefs is a data structure which represents what Ann believes Bob believes, and within this set P is included as well. Similarly included within this data structure is a data structure which represents what Ann believes Bob believes Ann believes (which also includes P), but now comes the trick: To get further nestings, we just include a pointer from this nested belief set back to the data structure of what Ann believes Bob believes. By following this pointer at the crucial point, we can get arbitrarily deep nestings, all within finite resources. Furthermore, the obvious implementation of this inclusion of data structures within data uses pointers uniformly to include structures within structures, so the fact that some of them point back up the hierarchy will not be noticed without special effort.

Here we have a line of inquiry that begins with the work of two philosophers and is applied independently to standard discourse by two psychologists and a computer scientist with the tools of computer science in this case the notions of data structures and recursive pointers, being used to finitely represent a seemingly infinite object, thereby giving additional credence to the psychological reality of this theory. Note, by the way, that while this data structure allows a system to reason to arbitrary depth about its beliefs, it cannot realize that it postulates common knowledge unless it can examine the entire structure globally and observe the loop in its representation.

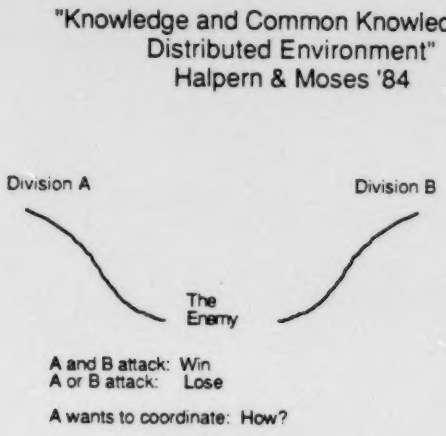


Figure 14.

Turning from discourse to pure computer science, we find some very exciting work, done fairly recently by Halpern and Moses that applies just these ideas to the notion of distributed computing, that argues that these ideas are central to getting processors to do things. Consider the following coordinated attack problem, which is discussed in the operating system literature (Fig. 14):

There are two divisions: Division A and Division B, each on its own hill, with a valley between them, and in the valley is The Enemy. Now, the commander of Division A realizes that if they both attack, they will wipe out The Enemy, but he also realizes that if only one of them attacks whoever attacks will be destroyed. How can he coordinate with Division B so that they will both attack? (This is before radio; the only way to communicate is by messenger.)

The problem is quite hard, if you think about it. If Commander A sends a messenger from A to B with the message "Attack at dawn, and I will too," if the messenger doesn't make it to B, then A attacks and gets destroyed. The obvious fix: B should send back a reply, saying the messenger got through; A only attacks if he hears back from B. But how does B know the messenger got through on the return trip? If she hasn't, then B attacks, and gets destroyed. It's a folk theorem of the operating system literature, that no matter how many messengers are used to verify that the last one got through, you still lose; there's no way to coordinate.

Halpern and Moses now have shown that by taking the notion of common knowledge discussed above, formalizing it, and applying the formalization to these kinds of problems, some extremely surprising theorems result.

First, they prove that there is no protocol of any kind for achieving common knowledge if communication is not guaranteed. From this it follows, they show, that common action cannot be achieved in a coordinated attack situation. Somewhat more precisely, Halpern and Moses show that any protocol that guarantees that if either party acts, then they both act, is a protocol in which necessarily neither party ever acts. Surprising. What happens if communication is guaranteed? (Maybe we can take our local network and make it really reliable...) Again, using their formalization of common knowledge, Halpern and Moses prove that even with guaranteed communications, if there is any uncertainty in each machine's clock, there is still no protocol for attaining common knowledge.

From this base, they consider synchronous broadcast and show that systems *can* achieve a weaker kind of common knowledge they call epsilon-common knowledge. Given that all processors know the (true) fact that epsilon is the propagation rate of the network, they show that at time two epsilon all processors can come to common knowledge that within time epsilon of when they hear a fact it will become common knowledge. They also talk about Eager Systems, where a processor in fact only knows P, but broadcasts to other processors that it is common knowledge that P. This message is overly eager and is false for a short time after transmission, at the least, but in return for the risks, it gets the job done much faster.

Back to the fundamental point, though. The real claim of the Halpern and Moses work is that the problems that come up in interactions between individuals are common to such interactions, whether the participants are machines or people, that notions of belief and mutual belief provide the underpinnings for any coordinated action or information flow.

A news flash: Within the last month, some important new ideas about mutual belief have been developed by Jon Barwise and Peter Aczel. Aczel is a mathematical logician working on set theories without a foundational axiom, in which there can be arbitrary infinite chains of descending sets. Playing with Aczel's logic, Barwise, a logician deeply interested in language, noticed a formalization of the notion of mutual belief which has a formal semantics. Let me use a simple example to illustrate the form of Barwise and Aczel's solution: Consider Bob and Ted playing a game of war, as shown in Figure 15, where Bob has the 3 of diamonds in front of him and Ted has the 4 of hearts in front of him. Within the larger situation that we see looking at Bob and Ted playing, there are two situations: Bob is in situation S1 in which he knows he has the 3 of diamonds, Ted has the 4 of hearts, and he also knows, since he can see Ted, that Ted sees situation S2 which is Ted's situation. What situation is S2? Well, what Ted sees, of course, is the same cards on the table, but he also sees that Bob sees

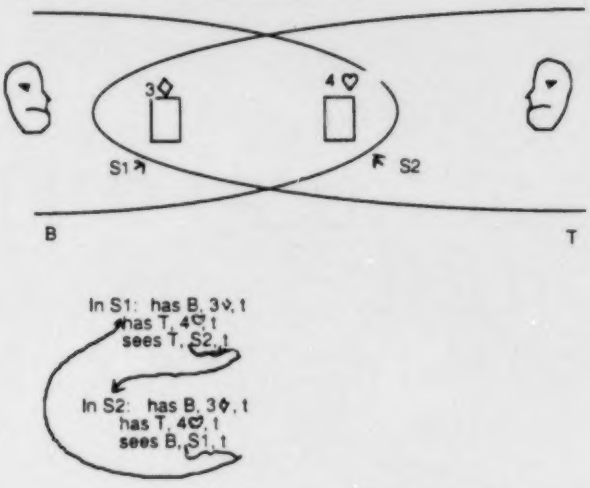


Figure 15.

situation S1. When one attempts to unravel the sets that correspond to these situations, one ends up with an infinite regress, but it is exactly this infinite regress that Aczel can handle and assign a semantics.

This formalization, by the way, begins to formalize the essence of Clark and Marshall's account of mutual belief, that mutual belief is often established by "copresence"; the idea that if we are both told something at the same time, and we both know we're both there, then we both mutually believe that we both know it. Barwise and Aczel's work begins to formalize what it means to be copresent in just this sense.

A brief mention of one last example, as a form of conclusion; the details of this example are too hard to explain briefly, but the idea is exciting with important implications. Brian Smith has been working on a language he calls 3-LISP. which is a semantically rationalized form of LISP (which is to say that it isn't LISP at all). Smith notes that any computational mechanism can be viewed as a single ball of fire, as it were, that does something, but it can also be viewed as an interpretive process, which means that it can be broken down into a plane of data structures, and a smaller ball of fire that runs over that set of data structures and does something which implements the larger ball of fire (Fig. 16). Smith then notes that this idea can be applied recursively. Take this second ball of fire and break it down into a set of data structures and an even smaller ball of fire. One can now decompose the interpreters through a potentially infinite regress, and 3-LISP is in fact designed to let the user do just this. Smith ties this idea

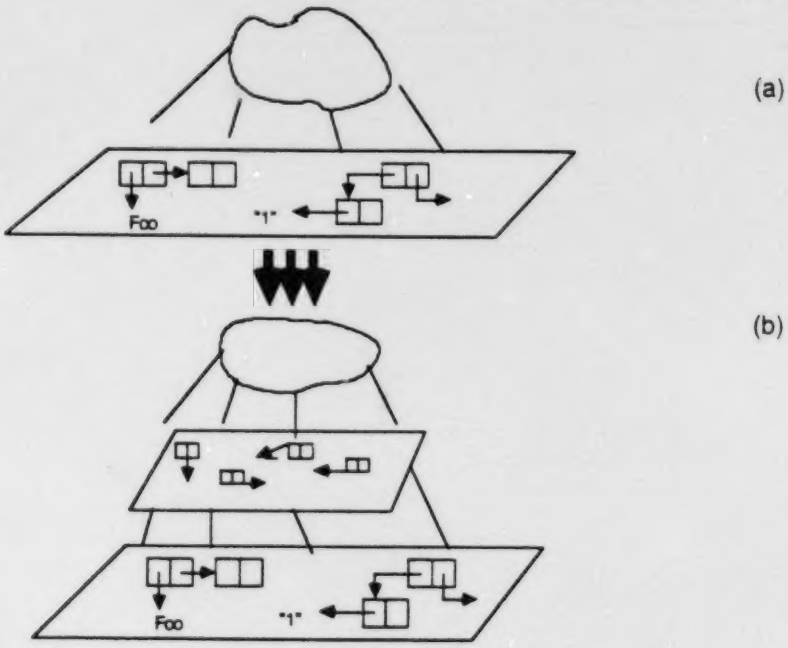


Figure 16.

to another: unlike most of the referring expressions in our head, which refer to things in the world about which we can't, therefore, actually *have* the referent, interpreters are special, in that the referring expressions in the implementation of an interpreter refer to entities which are in fact in the machine. Rather than rooting our formal semantic accounts on some particular machine that we happen to have, this level of abstracted data structures is a virtual machine which is going to be true of any instantiation of this language, as it were, because it is about the language.

One important implication of all this is that one can implement rich debugging environments that are both very powerful and very portable quite trivially in this kind of mechanism. Now I implemented the debugging hooks in one LISP implementation called Franz Lisp. This was done by writing the debugging hooks in C, because the LISP is itself implemented in C, but for certain operations, I needed to sneak my way past C into the machine stack itself, and then walking up the machine stack, decoding it as I go, so I can get pointers and smash them. Exactly how I did this depended heavily on the particulars of both the underlying computer language used to implement the LISP, and on the assembly language architecture of the computer on which the LISP was to be used; if either the underlying language or the machine

were changed, the implementation would have looked entirely different. This dependence of debugging tools on the particulars of the implementation is all rather common. The folks implementing 3-LISP, on the other hand, wrote a full powerful debugging environment in two days time; their implementation was done entirely within the semantic model of the virtual machine. So, the point here is that these ideas really have some interesting payoff. In fact, one bit of evidence that these ideas are really quite powerful is that they'll actually do work for us, and that provides important empirical evidence for the ideas which underlie them.

Acknowledgments

The editors of this report wish to express their heartfelt thanks to the staff and assistants who worked so many long hours to make it possible to prepare this document in a month's time. We especially want to thank: Kathy Adamczyk, Leslie Batema, David Brown, Arnold Chien, Ingrid Deiwiks, Dikran Karagueuzian, Jamie Marks, Purna Mishra, Emma Pease, Craige Roberts, Gautam Sengupta, Bach-Hong Tran.

—*Barbara Partee, Stanley Peters, Richmond Thomason*